09/896,896

Kongeriget Danmark

Patent application No.: PA 2000 01092

Date of filing: 14 July 2000

Applicant: Maxygen ApS
Agern Allé 1
DK-2970 Hørsholm

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims, sequence listing and figures as filed with the application on the filing date indicated above.

Applicant has changed its name from ProFound Pharma A/S to Maxygen ApS with effect from 01 September 2000. A transcript from the Danish Companies Register has been submitted to the Danish Patent Office as proof of the change of name. Consequently, Maxygen ApS is the same Applicant as ProFound Pharma A/S.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

Taastrup, 26 September 2002

Karin Schlichting
Head Clerk

PATENT- G VAREMÆRKESTYRELSEN

IMPROVED FOLLICLE STIMULATING HORMONE

FIELD OF THE INVENTION

The present invention relates to new polypeptides, to new polypeptide conjugates exhibiting follicle stimulating hormone (FSH) activity, to methods for preparing such polypeptides or conjugates, and to the use of such polypeptides or conjugates in therapy, in particular in the treatment of infertility.

BACKGROUND OF THE INVENTION

Follicle Stimulating Hormone (FSH) is a dimeric hormone consisting of an α subunit and a β subunit. The α subunit is common to the glycoprotein hormone family, which apart from FSH includes chorionic gonadotropin (GC), thyroid stimulating hormone (TSH), and luteinizing hormone (LH), whereas the β subunit is specific to FSH. The human wildtype α subunit is a 92 amino acid glycoprotein, the amino acid sequence of which is shown in SEQ ID NO 2. Said subunit is referred to herein as hFSH-α. The human wildtype β subunit is a 111 amino acid glycoprotein that has the amino acid shown in SEQ ID NO 4. This subunit is referred to herein as hFSH-β. hFSH-α comprises 5 cystines formed by the cysteines located in positions 7 and 31, 10 and 60, 28 and 82, 59 and 87, and 32 and 84, respectively. hFSH-β comprises 12 cysteines corresponding to 6 cystines located in positions 3 and 51, 17 and 66, 20 and 104, 28 and 82, 32 and 84, and 87 and 94, respectively.

Human FSH (hFSH) has been isolated from pituitary glands and from post-menopausal urine (EP 322 438) and has been produced recombinantly in mammalian cells (US 5,639,640, US 5,156,957, US 4,923,805, US 4,840,896, EP 211,894 and EP 521,586). The latter references also disclose the hFSH-β gene. US 5,405,945 discloses a modified human α subunit gene comprising only one intron.

US 4,589,402 and US 4,845,077 disclose purified hFSH which is free of LH and the use thereof for *in vitro* fertilization. EP 322 438 discloses a protein with at least 6200 U/mg FSH activity which is substantially free of LH activity, and wherein the FSH α subunit and β subunit, respectively, may be wildtype or specified truncated forms thereof.

Liu et al., J Biol Chem 1993, 15;268(2):21613-7, Grossmann et al., Mol Endocrinol 1996 10(6): 769-79, Roth and Dias (Mol Cell Endocribol 1995 1; 109(2): 143-9, Valove et al., Endochrinology 1994; 135(6):2657-61 ,Yoo et al., J Biol Chem1993 25; 268(18): 13034-42), US 5,508,261 and Chappel et al., 1998, Human Reproduction, 13(3): 18-35 disclose various structure-function relationship studies and identify amino acid residues involved in receptor binding and activation and in dimerization of FSH.

It has been found that glycosylation of FSH-α and FSH-β is essential for receptor signal transduction. hFSH-α comprises two N-glycosylation sites at the asparagines located at position 52 and 78, whereas hFSH-β comprises two N-glycosylation sites at the asparagines located at positions 7 and 24. The importance of the various N-glycosylation sites for the binding and signal-transducing activities of FSH are discussed, *inter alia*, by Valove et al., Endochrinology 1994; 135(6):2657-61 and Flack et al., J Biol Chem 1994 13;269(19):14015-20.

Galway et al., Endocrinology 1990; 127(1):93-100 demonstrate that FSH variants produced in a N-acetylglucosamine transferase-I CHO cell line or a CHO cell line defective in sialic acid

transport are as active as FSH secreted by wildtype cells or purified pituitary FSH *in vitro*, but lacked *in vivo* activity, possibly due to rapid clearance of the inadequately glycosylated variants in serum. D'Antonio et al., Human Reprod 1999; 14(5):1160-7 describe various FSH isoforms circulating in the blood stream. The isoforms have identical amino acid sequences, but differ in their extent of post-translational modification. It was found that the less acidic isoform group had a faster *in vivo* clearance as compared with the acidic isoform group, possibly due to differences in the sialic acid content between the isoforms. No significant difference in *in vitro* activity was observed between the isoforms. A similar result has been reported in US 5,087,615 and, for CHO produced recombinant FSH isoforms, by de Leeuw et al., Mol Hum Reprod 1996; 2(5):361-9.

US 5,087,615 discloses a method for stimulating follicle development and ovulation in a female patient by administering FSH to said patient during the follicular phase of the ovulatory cycle, the improvement comprising initially adminstering a first FSH isoform having a relatively long plasma half-life and subsequently administering a second FSH isoform having a shorter plasma half-life.

Bishop et al. Endochrinology 1995; 136(6):2635-40 conclude that circulatory half-life appears to be the primary determinant of *in vivo* activity.

Attempts have been made to prolong the serum half-life of FSH. US 5,338,835 and US 5,585,345 disclose a modified FSH-β subunit extended at the C-terminal Glu with the carboxy terminal portion (CTP) region of hCG (the entire region consisting of the amino acid sequence which occurs between positions 112-118 and 145, inclusive and comprising four O-linked glycosylation sites located at positions 121, 127, 132 and 138). The resulting modified subunit is stated to have the biological activity of native FSH, but a prolonged circulating half-life. US 5,405,945 discloses that the carboxy terminal portion of the CG β subunit or a variant thereof has significant effects on the clearance of GC, FSH, and LH.

US 5,883,073 discloses single-chain proteins comprised of two α-subunits with agonist or antagonist activity for CG, TSH, LH and FSH. The α subunits may be the human wildtype or a variant thereof, e.g. incorporating part of or the entire CTP region of hCG. Furthermore, the α subunit may be a variant in which amino acid residues between positions 50 and 60 are substituted, especially in positions 51, 53 and 55, or wherein Lys91 is converted to methionine or glutamic acid. The single-chain proteins can be combined with an appropriate β subunit.

US 5,508,261 discloses heterodimeric polypeptides having binding affinity to LH and FSH receptors comprising a glycoprotein hormone α subunit and a non-naturally occurring β subunit polypeptide, wherein the β subunit polypeptide is a chain of amino acids comprising four joined subsequences, each of which is selected from a list of specific sequences.

US 5,567,422 and WO 98/32466 suggest that FSH, among a vast number of other therapeutic proteins, may be PEGylated.

Currently, FSH is used therapeutically to stimulate the growth and maturation of ovarian follicles in infertile women. In particular, FSH is used in connection with *in vitro* fertilization as well as for the treatment of anovulatory women, with anovulatory syndrome or luteal phase deficiency. However, one problem encountered in current FSH treatment is the short *in vivo* half-life of FSH requiring frequent, usually daily administration of the product. The frequent administration is very inconvenient for the patient and results in high fluctuations of FSH activ-

ity in the blood stream, which is undesirable, and may cause inadequate maturation of the follicles.

Therefore, a clinical need exists for a product which provides part or all of the therapeutically relevant effects of FSH, and which may be administered at less frequent intervals as compared to currently available FSH product, and which preferably provides a more stable level of circulating FSH activity as compared to that obtainable by current treatment. The present invention is directed to such products as well as the means of making such products.

BRIEF DISCLOSURE OF THE INVENTION

More specifically, the present invention relates to polypeptide conjugates exhibiting FSH activity and methods for their preparation and their use in medical treatment.

Accordingly, in its first aspect the invention relates to a conjugate exhibiting FSH activity, comprising
i) a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits differs from the corresponding wildtype subunit in that at least one amino acid residue acid residue comprising an attachment group for a non-polypeptide moiety has been introduced or removed, and
ii) a non-polypeptide moiety bound to an attachment group of said polypeptide.

In a further aspect the invention relates to a polypeptide conjugate exhibiting FSH activity, comprising
i) a polypeptide comprising FSH-α and FSH-β subunits, wherein the amino acid sequence of at least one of said FSH-α and FSH-β subunits differs from that of the corresponding wildtype subunit in that at least one N-glycosylation site has been introduced, and
ii) an oligosaccharide moiety bound to an N-glycosylation site of said polypeptide.

In the above aspects the corresponding respective wildtype subunits are preferably hFSH-α and hFSH-β.

Another aspect of the invention relates to a polypeptide conjugate exhibiting FSH activity, comprising a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits comprises at least one introduced N- or O-glycosylation site at the N-terminal thereof, said at least one introduced glycosylation site being glycosylated.

In a further aspect, the invention relates to a polypeptide conjugate exhibiting FSH activity, comprising a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits comprises a polymer molecule bound to the N-terminal thereof.

In a still further aspect the invention relates to a substantially homogenous preparation of a conjugate of the invention.

In a further aspect the invention relates to generally novel modified FSH-α and modified FSH-β polypeptides. The polypeptides of the invention are contemplated to be useful as such for therapeutic, diagnostic or other purposes, but find particular interest as intermediate products for the preparation of a conjugate of the invention.

In still further aspects the invention relates to means and methods for preparing a conjugate or a polypeptide of the invention, including nucleotide sequences and expression vectors encoding a polypeptide or a conjugate of the invention.

In final aspects the invention relates to a therapeutic composition comprising a conjugate, polypeptide or preparation of the invention and methods of treating a mammal with such composition. In particular, the polypeptide, conjugate or composition of the invention may be used to treat infertility.

DETAILED DISCLOSURE OF THE INVENTION

Definitions

In the context of the present application and invention the following definitions apply:

The term "conjugate" is intended to indicate a heterogeneous molecule formed by the covalent attachment of one or more polypeptides to one or more non-polypeptide moieties such as polymer molecules, lipophilic compounds, carbohydrate moieties or organic derivatizing agents. The term covalent attachment means that the polypeptide and the non-polypeptide moiety are either directly covalently joined to one another, or else are indirectly covalently joined to one another through an intervening moiety or moieties, such as a bridge, spacer, or linkage moiety or moieties. Preferably, the conjugate is soluble at relevant concentrations and conditions, i.e. soluble in physiological fluids such as blood. The term "non-conjugated polypeptide" may be used about the polypeptide part of the conjugate.

The "polymer molecule" is a molecule formed by covalent linkage of two or more monomers, wherein none of the monomers is an amino acid residue, except where the polymer is human albumin or another abundant plasma protein. The term "polymer" may be used interchangeably with the term "polymer molecule". The term is intended to cover carbohydrate molecules attached by *in vitro* glycosylation. Carbohydrate molecules attached by *in vivo* glycolsylation, such as N- or O-glycosylation (as further described below) are referred to herein as "an oligosaccharide moiety". Except where the number of polymer molecules is expressly indicated, every reference to "a polymer", "a polymer molecule", "the polymer" or "the polymer molecule" contained in polypeptide of the invention or otherwise used in the present invention shall be a reference to one or more polymer molecule(s).

The term "attachment group" is intended to indicate a functional group of the polypeptide, in particular of an amino acid residue thereof or an oligosaccharide moiety, capable of attaching a non-peptide moiety such as a polymer molecule, a lipophilic molecule or an organic derivatizing agent. Useful attachment groups and their matching non-peptide moieties are apparent from the table below.

Attachment group	Amino acid	Examples of non-peptide moiety	Conjugation method/-Activated PEG	Reference
$-NH_2$	N-terminal, Lys	Polymer, e.g. PEG, with amide or imine group	mPEG-SPA Tresylated mPEG	Shearwater Inc. Delgado et al, critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992)
-COOH	C-term, Asp, Glu	Polymer, e.g. PEG, with ester or amide group Oligosaccharide moiety	mPEG-Hz *In vitro* coupling	Shearwater Inc.
-SH	Cys	Polymer, e.g. PEG, with disulfide, maleimide or vinyl sulfone group Oligosaccharide moiety	PEG-vinylsulphone PEG-maleimide *In vitro* coupling	Shearwater Inc. Delgado et al, critical reviews in Therapeutic Drug Carrier Systems 9(3,4):249-304 (1992)
-OH	Ser, Thr, OH-, Lys	Oligosaccharide moiety PEG with ester, ether, carbamate, carbonate	*In vivo* O-linked glycosylation	
$-CONH_2$	Asn as part of an N-glycosylation site	Oligosaccharide moiety Polymer, e.g. PEG	*In vivo N*-glycosylation	
Aromatic residue	Phe, Tyr, Trp	Oligosaccharide moiety	*In vitro* coupling	
$-CONH_2$	Gln	Oligosaccharide moiety	*In vitro* coupling	Yan and Wold, Biochemistry, 1984, Jul 31; 23(16): 3759-65
Aldehyde Ketone	Oxidized oligo-saccharide	Polymer, e.g. PEG, PEG-hydrazide	PEGylation	Andresz et al., 1978, Makromol. Chem. 179:301, WO 92/16555, WO 00/23114

Guanidino	Arg	Oligosaccharide moiety	*In vitro* coupling	Lundblad and Noyes, Chimical Reagents for Protein Modification, CRC Press Inc. Boca Raton, Fl
Imidazole ring	His	Oligosaccharide moiety	*In vitro* coupling	As for guanidine

For *in vivo* N-glycosylation, the term "attachment group" is used in an unconventional way to indicate the amino acid residues constituting an N-glycosylation site (with the sequence N-X'-S/T/C-X'', wherein X' is any amino acid residue except proline, X'' any amino acid residue which may or may not be identical to X' and which preferably is different from proline, N is asparagine, and S/T/C is either serine, threonine or cysteine, preferably serine or threonine, and most preferably threonine). Although the asparagine residue of the N-glycosylation site is where the oligosaccharide moiety is attached during glycosylation, such attachment cannot be achieved unless the other amino acid residues of the N-glycosylation site are present. Accordingly, when the non-peptide moiety is an oligosaccharide moiety and the conjugation is to be achieved by N-glycosylation, the term "amino acid residue comprising an attachment group for the non-peptide moiety" as used in connection with alterations of the amino acid sequence of the polypeptide of interest is to be understood as meaning that one or more amino acid residues constituting an N-glycosylation site are to be altered in such a manner that either a functional N-glycosylation site is introduced into the amino acid sequence or removed from said sequence.

In the present application, amino acid names and atom names (e.g. CA, CB, NZ, N, O, C, etc.) are used as defined by the Protein DataBank (PDB), which is based on the IUPAC nomenclature (IUPAC Nomenclature and Symbolism for Amino Acids and Peptides (residue names, atom names etc.), *Eur. J. Biochem.*, 138, 9-37 (1984) together with their corrections in *Eur. J. Biochem.*, 152, 1 (1985). The term "amino acid residue" is intended to indicate an amino acid residue contained in the group consisting of alanine (Ala or A), cysteine (Cys or C), aspartic acid (Asp or D), glutamic acid (Glu or E), phenylalanine (Phe or F), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), lysine (Lys or K), leucine (Leu or L), methionine (Met or M), asparagine (Asn or N), proline (Pro or P), glutamine (Gln or Q), arginine (Arg or R), serine (Ser or S), threonine (Thr or T), valine (Val or V), tryptophan (Trp or W), and tyrosine (Tyr or Y) residues. The terminology used for identifying amino acid positions/substitutions is illustrated as follows: E9(a) indicates position #9 occupied by a glutamic acid residue in the amino acid sequence shown in SEQ ID NO 2. E9(a)N indicates that said glutamic acid residue has been substituted by an asparagine residue. The numbering of amino acid residues made herein is made relative to the amino acid sequence shown in SEQ ID NO 2 (for FSH-α) and SEQ ID NO 4 (for FSH-β). Multiple substitutions are indicated with a "+", e.g. M109(b)N+E111(b)S/T means an amino acid sequence which comprises a substitution of the methionine residue in position 109 of FSH-β by an asparagine residue and a substitution of the glutamic acid residue in position 111 in FSH-β by a serine or a threonine residue.

The term "nucleotide sequence" is intended to indicate a consecutive stretch of two or more nucleotide molecules. The nucleotide sequence may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combination thereof.

The term "polymerase chain reaction" or "PCR" generally refers to a method for amplification of a desired nucleotide sequence in vitro, as described, for example, in US 4,683,195. In general, the PCR method involves repeated cycles of primer extension synthesis, using oligonucleotide primers capable of hybridising preferentially to a template nucleic acid.

"Cell", "host cell", "cell line" and "cell culture" are used interchangeably herein and all such terms should be understood to include progeny resulting from growth or culturing of a cell. "Transformation" and "transfection" are used interchangeably to refer to the process of introducing DNA into a cell.

"Operably linked" refers to the covalent joining of two or more nucleotide sequences, by means of enzymatic ligation or otherwise, in a configuration relative to one another such that the normal function of the sequences can be performed. For example, the nucleotide sequence encoding a presequence or secretory leader is operably linked to a nucleotide sequence for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide: a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the nucleotide sequences being linked are contiguous and, in the case of a secretory leader, contiguous and in reading phase. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, then synthetic oligonucleotide adaptors or linkers are used, in conjunction with standard recombinant DNA methods.

The term "introduce" is primarily intended to mean substitution of an existing amino acid residue, but may also mean insertion of an additional amino acid residue. The term "remove" is primarily intended to mean substitution of the amino acid residue to be removed by another amino acid residue, but may also mean deletion (without substitution) of the amino acid residue to be removed.

The term "immunogenicity" as used in connection with a given substance is intended to indicate the ability of the substance to induce a response from the immune system. The immune response may be a cell or antibody mediated response (see, e.g., Roitt: Essential Immunology (8th Edition, Blackwell) for further definition of immunogenicity). Normally reduced antibody reactivity will be an indication of a reduced immunogenicity.

The term "functional *in vivo* half-life" is used in its normal meaning, i.e. the time at which 50% of the biological activity of the polypeptide or conjugate is still present in the body/target organ, or the time at which the activity of the polypeptide or conjugate is 50% of the initial value. As an alternative to determining functional *in vivo* half-life, "serum half-life" may be determined, i.e. the time at which 50% of the dispensed polypeptide or conjugate molecules is still present in the circulation/plasma/bloodstream. Determination of serum half-life is often more simple than determining the functional *in vivo* half-life and the magnitude of serum half-life is usually a good indication of the magnitude of functional *in vivo* half-life. Alternative terms to serum half-life include "plasma half-life", "circulating half-life", "serum clearance", "plasma clearance" and "clearance half-life". The polypeptide or conjugate is cleared by the action of one or more of the kidney, reticuloendothelial systems (RES), spleen or liver, by FSH-receptor-mediated elimination, or by specific or non-specific proteolysis. Normally, clearance depends on size (relative to the cutoff for glomerular filtration), charge, attached carbohydrate chains, and the presence of cellular receptors for the protein. The functionality to be retained is normally selected from proliferative or receptor binding

activity. The functional *in vivo* half-life and the serum half-life may be determined by any suitable method known in the art as further discussed in the Materials and Methods section hereinafter.

The term "increased" as used about the functional *in vivo* half-life or serum half-life is used to indicate that the relevant half-life of the conjugate or polypeptide is statistically significantly increased relative to that of a reference molecule, such as a non-conjugated rhFSH (recombinant hFSH), e.g. Gonal-F® (available from Serono) or Puregon® (available from Organon), as determined under comparable conditions.

The term "renal clearance" is used in its normal meaning to indicate any clearance taking place by the kidneys, e.g. by glomerular filtration, tubular excretion or degradation in the tubular cells. Renal clearance depends on physical characteristics of the conjugate, including size (diameter), symmetry, shape/rigidity and charge. A molecular weight of about 67 kDa is considered to be an important cut-off-value for renal clearance, i.e. a molecular weight above about 67 kDa normally results in reduced renal clearance. A reduced renal clearance may be confirmed by any suitable assay, e.g. an established *in vivo* assay. Typically, the renal clearance is determined by administering a labelled (e.g. radiolabelled or fluorescenctly labelled) polypeptide conjugate to a patient and measuring the label activity in urine collected from the patient during a specified time. The reduced renal clearance is determined relative to the corresponding non-conjugated polypeptide or the non-conjugated corresponding wild-type polypeptide under comparable conditions.

The term "FSH-α" is intended to indicate a polypeptide having qualitatively similar functions or activities as the corresponding wildtype FSH α subunit, including the capability of forming a dimeric polypeptide with an FSH-β subunit (FSH-β), which dimeric polypeptide exhibits FSH activity. Alternatively used terms include "FSH-α polypeptide", "FSH-α subunit", and "modified FSH-α". Analogously, the term "FSH-β" is intended to indicate a polypeptide having qualitatively similar functions or activities as the corresponding wildtype FSH β subunit, including the capability of dimerizing with FSH-α and thereby forming a dimeric polypeptide exhibiting FSH activity. Alternatively used terms include "FSH-β polypeptide", "FSH-β subunit", and "modified FSH-β".

The term "exhibiting FSH activity" is intended to indicate that the conjugate or polypeptide has one or more of the functions of wildtype FSH, in particular hFSH, including the capability of binding to and activating a FSH receptor. The FSH activity is conveniently assayed using the receptor binding assay described in the Materials and Methods section hereinafter. The conjugate or polypeptide "exhibiting" FSH activity is considered to have such activity when it displays a measurable function, e.g. a measurable activity. The dimeric polypeptide exhibiting FSH activity may also be termed " FSH molecule" herein.

Conjugate of the invention

As stated above, in a first aspect the invention relates to a polypeptide conjugate exhibiting FSH activity, comprising i) a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of the FSH-α and FSH-β subunits differs from the corresponding wildtype subunit in at least one introduced or removed amino acid residue comprising an attachment group for non-polypeptide moiety, and ii) a non-polypeptide moiety bound to an attachment group of the polypeptide. Examples of amino acid residues that may be introduced and/or removed are described in further detail in the following sections.

The conjugate of the invention is the result of a generally new strategy for developing improved molecules with FSH activity. More specifically, by removing and/or introducing an amino acid residue comprising an attachment group for the non-polypeptide moiety it is possible to specifically adapt the polypeptide so as to make the molecule more susceptible to conjugation to the non-polypeptide moiety of choice, to optimize the conjugation pattern (e.g. to ensure an optimal distribution of non-polypeptide moieties on the surface of the FSH molecule and to ensure that only the attachment groups intended to be conjugated are present in the molecule) and thereby obtain a new conjugate molecule which has FSH activity and in addition one or more improved properties as compared to FSH molecules available today, in particular increased functional *in vivo* half-life and/or reduced renal clearance.

In the conjugate of the invention, one or both of the FSH subunits may be modified according to the invention. For instance, the amino acid sequence of FSH-α may be modified as described herein, whereas FSH-β is unmodified, and vice versa. Alternatively, both of FSH-α and FSH-β may be modified according to the invention.

While the FSH-α and/or FSH-β may be of any origin, in particular mammalian origin, it is presently preferred that they are of of human origin. Accordingly, the corresponding wildtype subunits referred to above are preferably hFSH-α and hFSH-β, respectively, with the amino acid sequences shown in SEQ ID NO 2 and 4, respectively.

In a preferred embodiment one difference between the amino acid sequence of FSH-α and/or FSH-β and the corresponding wildtype sequence is that at least one and preferably more, e.g. 1-15, amino acid residues comprising an attachment group for the non-polypeptide moiety ii) have been introduced, preferably by substitution, into the amino acid sequence(s). Thereby, for instance, shielding by non-polypeptide moieties may be achieved in different regions of the polypeptide molecule, leading to a lower immune response, and/or the molecular weight, shape, size and/or charge of the conjugate can be optimised. Preferably, such amino acid residues are introduced in positions occupied by an amino acid residue having more than 25%, such as more than 50% or even more than 75% of its side chain exposed at the surface of the molecule.

The term "one difference" as used in the present application is intended to allow for additional differences being present. Accordingly, in addition to the specified amino acid difference, other amino acid residues than those specified may be mutated.

In a further preferred embodiment one difference between the amino acid sequence of FSH-α and/or FSH-β and that of the corresponding wildtype polypeptide is that at least one and preferably more, e.g. 1-15, amino acid residues comprising an attachment group for the non-polypeptide moiety ii) have been removed, preferably by substitution, from the amino acid sequence. The amino acid residue to be removed is preferably one to which conjugation is disadvantageous, e.g. an amino acid residue located at or near a functional site of the polypeptide (since conjugation at such a site may result in inactivation or reduced FSH activity of the resulting conjugate due to impaired receptor recognition). In the present context the term "functional site" is intended to indicate one or more amino acid residues which are essential for or otherwise involved in the function or performance of hFSH, in particular dimerization and/or receptor binding and activation. Such amino acid residues are a part of a functional site. The functional site may be determined by methods known in the art and is preferably identified by analysis of a structure of the polypeptide complexed to a relevant receptor, such as the hFSH receptor.

In preferred embodiments of the present invention more than one amino acid residue of the FSH-α and/or FSH-β is altered, e.g. the alteration embraces removal as well as introduction of amino acid residues comprising an attachment group for the non-polypeptide moiety of choice.

Typically, in order to avoid too much disruption of the structure and function of the FSH molecule the total number of amino acid residues to be altered in accordance with the present invention does not exceed 15. Preferably, the polypeptide part of the conjugate of the invention or the polypeptide of the invention comprises an amino acid sequence which differs in 1-15 amino acid residues from the amino acid sequence shown in SEQ ID NO 2, such as in 1-8 or 2-8 amino acid residues, e.g. in 1-5 or 2-5 amino acid residue from the amino acid sequence shown in SEQ ID NO 2. Thus, normally the polypeptide part of the conjugate or the polypeptide of the invention comprises an amino acid sequence which differs from the amino acid sequence shown in SEQ ID NO 2 in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 amino acid residues.

The FSH-α and/or FSH-β of the polypeptide i) is preferably any of the specific modified FSH-α and/or FSH-β polypeptides disclosed in the subsequent sections having introduced and/or removed amino acid residues comprising an attachment group for the relevant non-polypeptide moiety.

The amino acid residue comprising an attachment group for a non-polypeptide moiety, whether it is removed or introduced, is selected on the basis of the nature of the non-polypeptide moiety of choice and, in most instances, on the basis of the method in which conjugation between the polypeptide i) and the non-polypeptide moeity ii) is to be achieved. It will be understood that in order to preserve a measurable function of the modified FSH-α and/or FSH-β, amino acid residues to be modified (by deletion, preferably by substitution) are selected from those amino acid residues which are not essential for providing a measurable activity. Accordingly, amino acid residues to be modified are different from those required for subunit dimerization and/or receptor binding or activation. The identity of such amino acid residues is described in the prior art (a representative part of which is identified in the Background section above) or can be determined by a person skilled in the art using methods known in the art.

In addition to the removal and/or introduction of amino acid residues the FSH-α and/or FSH-β may comprise other amino acid changes, such as substitutions, or glycosylations which are not related to introduction and/or removal of amino acid residues comprising an attachment group for the non-polypeptide moiety. Examples of such additional amino acid changes include adding part of or the entire CTP region of hGC to the C-terminus of FSH-α or introducing any other mutation (in particular selected among those reported to enhance FSH activity and/or increase the functional *in vivo* half-life, cf. the Background of the Invention section herein.)

Preferably, the conjugate of the present invention has one or more improved properties as compared to hFSH, including increased functional *in vivo* half-life, increased serum half-life, reduced renal clearance, reduced immunogenicity and/or an increased bioavailability as compared to rhFSH (e.g. Gonal-F® or Puregon®). Consequently, medical treatment with a conjugate of the invention offers a number of advantages over the currently available FSH compounds, including longer duration between injections and fewer side effects.

Normally, the increased functional *in vivo* half-life is obtained as a consequence of the conjugate having a reduced susceptibility to renal clearance as compared to hFSH. The reduced susceptibility to renal clearance is obtained as a consequence of the size, shape/rigidity, net charge and other characteristics of the conjugate being changed as compared to the unconjugated polypeptide. In a preferred embodiment, the conjugate according to the invention has a molecular weight of at least about 67 kDa, preferably at least about 70 kDa, although a lower molecular weight may also give rise to a reduced renal clearance. In some cases, it will be preferred to obtain a slightly reduced renal clearance, e.g. to increase the *in vivo* half-life from about 24 hours to about 3-4 days, but to avoid a longer half-life of e.g. about a week. In such cases, the conjugate of the invention may have a molecular weight that is substantially below about 67 kDa, but which nevertheless has been increased a sufficient amount so as to ensure a desired reduction in renal clearance. Polymer molecules, such as PEG, have been found to be particularly useful for adjusting the molecular weight of the conjugate. As will be explained in further detail below, the number and size of such polymer molecules may be adapted in order to obtain a desired renal clearance, as well as other desired properties, suitable for a given clinical indication.

In a preferred embodiment, the conjugate of the invention has a reduced renal clearance of at least about 50%, such as least about 75% or at least about 90%, as compared to the corresponding non-conjugated polypeptide (such as hFSH or rhFSH) as determined under comparable conditions.

Conjugate of the invention wherein the non-polypeptide moiety is attached to a lysine or the N-terminal amino acid residue

In a preferred embodiment the conjugate of the invention is one wherein the amino acid residue comprising an attachment group for the non-polypeptide moiety is a lysine residue and the non-polypeptide moiety ii) is any molecule which has lysine as an attachment group. For instance, the non-polypeptide moiety may be a polymer molecule, in particular any of the molecules mentioned in the section entitled "Conjugation to a polymer molecule", and preferably selected from the group consisting of linear or branched polyethylene glycol and polyalkylene oxide. Most preferably, the polymer molecule is mPEG-SPA or oxycarbonyl-oxy-N-dicarboxyimide PEG (US 5,122,614).

The FSH-α and/or FSH-β having introduced and/or removed at least one lysine may advantageously be *in vivo* glycosylated, e.g. using naturally occurring glycosylation sites present in the relevant FSH polypeptide. However, in a particular embodiment the conjugate is one wherein the amino acid sequence of FSH-α and/or FSH β differs from that of FSH-α and/or FSH-β in that an N-glycosylation site has been introduced and/or removed. Such introduced/removed sites may be any of those described in the section entitled "Conjugate of the invention wherein the non-polypeptide moiety is an oligosaccharide moiety".

i) Removal of lysine residues

hFSH-α contains 6 lysine residues and hFSH-β 7. In order to avoid conjugation to one or more of these lysine residues, e.g. lysine residues located at or close to the receptor-binding site of hFSH, it may be desirable to remove at least one lysine residue. Accordingly, in one embodiment the conjugate of the invention is one which comprises a modified FSH-α having an amino acid residue which differs from that of hFSH-α in the removal of at least one lysine residue selected from the group consisting of K44(a), K45(a), K51(a), K63(a), K75(a), and

K91(a), in particular at least one amino acid residue selected from of the group consisting of K44(a), K45(a), K63(a), K75(a), and K91(a) (these residues having more than 25% of their side chain exposed to the surface), and preferably from the group consisting of K45(a), K63(a), K75(a), and K91(a) (these residues having more than 50% of their side chain exposed to the surface). The FSH-β part of this conjugate may be hFSH-β or any of the modified FSH-β polypeptides described herein.

In another embodiment the conjugate of the invention is one which comprises a modified FSH-β having an amino acid residue which differs from that of hFSH-β in the removal of at least one lysine residue selected from the group consisting of K14(b), K40(b), K46(b), K49(b), K54(b), K86(b), and K110(b), in particular at least one amino acid residue selected from of the group consisting of K14(b), K40(b), K46(b), K49(b), K54(b), K86(b), and K110(b) (these residues having more than 25% of their side chain exposed to the surface), and preferably from the group consisting of K46(b), K54(b), K86(b), and K110(b) (these residues having more than 50% of their side chain exposed to the surface). The FSH-α part of this conjugate may be hFSH-α or any of the modified FSH-α polypeptides described herein.

In a further embodiment, the conjugate of the invention is one which comprises a modified FSH-α and a modified FSH-β, each of which differ from the corresponding hFSH subunit in the removal of at least one of the above identified lysine residues. For instance, the conjugate of the invention may be one wherein the modified FSH-α and modified FSH-β subunit differ from the corresponding hFSH subunit in at least one of K45(a), K63(a), K75(a), and K91(a) and at least one of K46(b), K54(b), K86(b), and K110(b).

The removal of any of the above lysine residues is preferably achieved by substitution by any other amino acid residue, in particular by an arginine or a glutamine residue.

ii) Introduction of lysine residues

In order to obtain a more extensive conjugation it may be desirable to introduce at least one non-naturally occurring lysine residue in hFSH, in particular in a position occupied by an amino acid residue having a side chain which is more than 25% surface exposed and which is not part of a cystine or located at a receptor binding site. Such amino acid residues are identified in the Examples section hereinafter or form part of the state of the art.

Accordingly, in a further embodiment the conjugate of the invention is one which comprises a modifed FSH-α having an amino acid residue which differs from that of hFSH-α in the introduction of at least one lysine residue in a position selected from the group consisting of A1(a), P2(a), D3(a), V4(a), Q5(a), D6(a), P8(a), E9(a), T11(a), L12(a), Q13(a), E14(a), P16(a), F17(a), Q20(a), P21(a), G22(a), A23(a), P24(a), L26(a), M29(a), F33(a), R42(a), S43(a), T46(a), L48(a), V49(a), Q50(a), N52(a), V61(a), S64(a), Y65(a), N66(a), R67(a), V68(a), T69(a), M71(a), G72(a), G73(a), F74(a), N78(a), T80(a), A81(a), H83(a), S85(a), T86(a), Y88(a), Y89(a), H90(a), and S92(a), in particular selected from of the group consisting of A1(a), P2(a), D3(a), V4(a), Q5(a), D6(a), P8(a), E9(a), T11(a), Q13(a), E14(a), P16(a), F17(a), Q20(a), P21(a), G22(a), A23(a), T46(a), L48(a), V49(a), Q50(a), N52(a), S64(a), N66(a), R67(a), T69(a), G72(a), G73(a), T86(a), Y89(a), H90(a), and S92(a) (these residues having more than 50% of their side chain exposed to the surface), and most preferably in the position R42(a) and/or R67(a), such as R67(a). The FSH-β part of this conjugate may be hFSH-β or any of the modified FSH-β polypeptides described herein.

In a further embodiment the conjugate of the invention is one which comprises a modifed FSH-β having an amino acid residue which differs from that of hFSH-β in the introduction of at least one lysine residue in a position selected from the group consisting of N1(b), S2(b), E4(b), L5(b), T6(b), N7(b), I8(b), T9(b), E15(b), E16(b), R18(b), F19(b), I21(b), S22(b), N24(b), Y31(b), Y33(b), R35(b), D36(b), L37(b), Y39(b), D41(b), P42(b), A43(b), R44(b), P45(b), I47(b), E55(b), L56(b), V57(b), Y58(b), E59(b), T60(b), V61(b), R62(b), P64(b), G65(b), A67(b), H68(b), H69(b), D71(b), L73(b), Y74(b), T75(b), T80(b), Q81(b), H83(b), G85(b), D88(b), S89(b), D90(b), S91(b), D93(b), T95(b), V96(b), R97(b), G98(b), L99(b), G100(b), Y103(b), S105(b), F106(b), G107(b), E108(b), M109(b), and E111(b), in particular selected from of the group consisting of N1(b), N7(b), T9(b), E15(b), E16(b), R18(b), F19(b), N24(b), Y33(b), D41(b), P42(b), A43(b), R44(b), P45(b), I47(b), E55(b), V57(b), Y58(b), E59(b), R62(b), P64(b), G65(b), A67(b), H68(b), H69(b), D71(b), L73(b), T75(b), Q81(b), H83(b), D88(b), S89(b), D90(b), S91(b), T95(b), R97(b), G98(b), L99(b), G100(b), Y103(b), S105(b), F106(b), G107(b), E108(b), M109(b), and E111(b) (these residues having more than 50% of their side chain exposed to the surface), and most preferably selected from the group consisting of R18(b), R35(b), R44(b), R62(b), and R97(b), such R18(b), R44(b), R62(b), and R97(b). The FSH-α part of this conjugate may be hFSH-α or any of the modified FSH-α polypeptides described herein.

In a further embodiment, the conjugate of the invention is one which comprises a modified FSH-α and a modified FSH-β, each of which differ from the corresponding hFSH subunit in the introduction of a lysine residue in at least one of the above identified positions. For instance, the conjugate of the invention may be one wherein the modified FSH-α and modified FSH-β subunit differ from the corresponding hFSH subunit in that a lysine residue has been introduced in at least one of R42(a) and R67(a), and at least one of R18(b), R35(b), R44(b), R62(b), and R97(b), and more preferably in R67(a), and at least one of R18(b), R44(b), R62(b), R97(b).

The introduction of a lysine residue is preferably achieved by substitution of any of the above amino acid residues.

iii) Introduction and removal of lysine residues

In a preferred embodiment the conjugate of the invention comprises at least one introduced lysine residue, in particular any of those described in the section entitled "Introduction of lysine residues", and at least one removed lysine residue, in particular any of those described in the section entitled "Removal of lysine residues".

Preferably, the conjugate comprises a modified FSH-α and/or a modified FSH-β which differs from the corresponding hFSH-α/β in at least one introduced and at least one removed lysine residue, wherein the lysine residue is introduced by substitution of an amino acid residue selected from the group consisting of R42(a) and R67(a), R18(b), R35(b), R44(b), R62(b), and R97(b), and more preferably from the group consisting of R67(a), R18(b), R44(b), R62(b), and R97(b) and removal of a lysine residue selected from the group consisting of K45(a), K63(a), K75(a), K91(a) K46(b), K54(b), K86(b), and K110(b), the removal preferably being achieved by substitution by any other amino acid residue, in particular by an arginine residue.

N-terminal PEGylation of FSH

As indicated above, one aspect of the invention relates to a polypeptide conjugate wherein at least one of the FSH-α and FSH-β subunits comprises a polymer molecule bound to the N-

terminal thereof. Preferably, the polymer is a polyethylene glycol (PEG) such as mPEG; see the general discussion below regarding conjugates comprising polyethylene glycol-derived polymers.

In the case of N-terminal PEGylated FSH conjugates according to the invention, the respective subunits may comprise one or more of the modifications disclosed elsewhere herein, or one or both of the subunits may be the respective wildtype subunits with a PEG-derived polymer being attached at the N-terminal. Thus, the polypeptide conjugate may be one in which the FSH-α subunit comprises hFSH-α having the sequence shown in SEQ ID NO 2, and/or in which the FSH-β subunit comprises hFSH-β having the sequence shown in SEQ ID NO 4. In a particular embodiment, both of the subunits correspond to the respective wildtype hFSH subunits, although with either the α or β subunit, or both, being N-terminally PEGylated.

Aldehyde-activated PEG and reduction using $NaBH_3CN$ have been used to selectively pegylate the N-terminal α-amino group of proteins (see for instance US 5,824,784 regarding N-terminal PEGylation of G-CSF). The N-terminus of the α and/or the β chain of wildtype FSH or a modified form of FSH can be PEGylated using similar methods. Reaction materials include purified FSH or a modified form of FSH, methoxy-PEG-aldehyde (M-PEG-CHO), and $NaBH_3CN$. In order to optimise yield, one may for instance vary: molar ratio of FSH, M-PEG-CHO and $NaBH_3CN$, time for establishment of the Schiff's base equilibrium (reaction between FSH and M-PEG-CHO before addition of $NaBH_3CN$), reaction time after addition of $NaBH_3CN$, temperature, pH, or reaction volume. The yield of PEGylated FSH forms may be analysed using Western blotting, mass spectrometry and N-terminal sequencing. In order to restrict PEGylation to only one of the two N-termini in FSH, PEGylation of the α or β chain may be selectively prevented by addition of a glutamine to the N-terminus. Spontaneous cyclisation of such an N-terminal glutamine residue will render it unaccessible for PEGylation. Such a glutamine residue may subsequently be removed using a pyroglutamyl aminopeptidase (for instance EC 3.4.19.3).

Conjugate of the invention having a non-lysine residue as an attachment group

Based on the present disclosure the skilled person will be aware that amino acid residues comprising other attachment groups may be introduced into and/or removed from FSH-α and/or FSH-β, using the same approach as that illustrated above by lysine residues. For instance, one or more amino acid residues comprising an acid group (glutamic acid and aspartic acid), asparagine, tyrosine and cysteine may be introduced into positions which in hFSH are occupied by amino acid residues having surface exposed side chains (i.e. the positions mentioned above as being of interest for introduction of lysine residues), or removed (preferably by substitution by any other amino acid residue). Preferably, Asp is substituted by Asn, Glu by Gln, Tyr by Phe, and Cys by Ser.

Conjugate of the invention wherein the non-polypeptide moiety is an oligosaccharide moiety

It has been found that N-glycosylation is important for FSH activity and also that the extent and type of oligosaccharide moiety attached by *in vivo* glycosylation is important for functional *in vivo* half-life of the glycosylated FSH. In order to obtain a different, optionally increased glycosylation it is desirable to introduce at least one glycosylation site. Accordingly, in a further aspect the invention relates to polypeptide conjugate exhibiting FSH activity comprising i) a polypeptide comprising FSH-α and FSH-β, wherein the amino acid sequence of said FSH-α and/or FSH-β differs from that of the corresponding wild type FSH, preferably hFSH, in at least one introduced N-glycosylation site and ii) an oligosaccharide moiety.

A suitable N-glycosylation site may be introduced by introducing, preferably by substitution, an asparagine residue in a position occupied by an amino acid residue having more than 25% of its side chain exposed at the surface of the polypeptide, which position does not have a proline residue located in position +1 or +3 therefrom. If the amino acid residue located in position +2 is a serine or threonine, no further amino acid substitution is required. However, if this position is occupied by a different amino acid residue, a serine or threonine residue needs to be introduced.

Preferably, the conjugate according to this embodiment is one which comprises a modified FSH-α having an amino acid residue which differs from that of hFSH-α in the introduction of at least one N-glycosylation site by a mutation selected from the group consisting of P2(a)N+V4(a)S, P2(a)N+V4(a)T, D3(a)N+Q5(a)S, D3(a)N+Q5(a)T, V4(a)N+D6(a)S, V4(a)N+D6(a)S, D6(a)N+P8(a)S, D6(a)N+P8(a)T, E9(a)N+T11(a)S, E9(a)N, T11(a)N+Q13(a)S, T11(a)N+Q13(a)T, L12(a)N+E14(a)S, L12(a)N+E14(a)T, E14(a)N+P16(a)S, E14(a)N+P16(a)T, P16(a)N+F18(a)S, P16(a)N+F18(a)T, F17(a)N, F17(a)N+S19(a)T, G22(a)N+P24(a)S, G22(a)N+P24(a)T, P24(a)N+L26(a)S, P24(a)N+L26(a)T, F33(a)N+R35(a)S, F33(a)N+R35(a)T, R42(a)N+K44(a)S, R42(a)N+K44(a)T, S43(a)N+K45(a)S, S43(a)N+K45(a)T, K44(a)N+T46(a)S, K44(a)N, K45(a)N+M47(a)S, K45(a)N+M47(a)T, T46(a)N+L48(a)S, T46(a)N+L48(a)T, L48(a)N+Q50(a)S, I48(a)N+Q50(a)T, V49(a)N+K51(a)S, V49(a)N+K51(a)T, Q50(a)N+N52(a)S, Q50(a)N+N52(a)T, V61(a)N+K63(a)S, V61(a)N+K63(a)T, K63(a)N+Y65(a)S, K63(a)N+Y65(a)T, S64(a)N+N66(a)S, S64(a)N+N66(a)T, Y65(a)N+R67(a)S, Y65(a)N+R67(a)T, V68(a)S, V68(a)T, R67(a)N+T69(a)S, R67(a)N, T69(a)N+M71(a)S, T69(a)N+M71(a)T, M71(a)N+G73(a)S, M71(a)N+G73(a)T, G72(a)N+F74(a)S, G72(a)N+F74(a)T, G73(a)N+K75(a)S, G73(a)N+K75(a)T, F74(a)N+V76(a)S, F74(a)N+V76(a)T, K75(a)N+E77(a)S, K75(a)N+E77(a)T, A81(a)N+H83(a)S, A81(a)N+H83(a)T, H83(a)N, T86(a)N+Y88(a)S, T86(a)N+Y88(a)T, Y88(a)N+H90(a)S, Y88(a)N+H90(a)T, Y89(a)N+K91(a)S, Y89(a)N+K91(a)T, H90(a)N and H90(a)N+S92(a)T, more preferably from the group consisting of V68(a)S, V68(a)T, E9(a)N, F17(a)N, K44(a)N, R67(a)N, H83(a)N and H90(a)N, even more preferably from the group consisting of P2(a)N+V4(a)S, P2(a)N+V4(a)T, D3(a)N+Q5(a)S, D3(a)N+Q5(a)T, V4(a)N+D6(a)S, V4(a)N+D6(a)S, D6(a)N+P8(a)S, D6(a)N+P8(a)T, E9(a)N+T11(a)S, E9(a)N, T11(a)N+Q13(a)S, T11(a)N+Q13(a)T, E14(a)N+P16(a)S, E14(a)N+P16(a)T, P16(a)N+F18(a)S, P16(a)N+F18(a)T, F17(a)N, F17(a)N+S19(a)T, G22(a)N+P24(a)S, G22(a)N+P24(a)T, K45(a)N+M47(a)S, K45(a)N+M47(a)T, T46(a)N+L48(a)S, T46(a)N+L48(a)T, L48(a)N+Q50(a)S, I48(a)N+Q50(a)T, V49(a)N+K51(a)S, V49(a)N+K51(a)T, Q50(a)N+N52(a)S, Q50(a)N+N52(a)T, K63(a)N+Y65(a)S, K63(a)N+Y65(a)T, S64(a)N+N66(a)S, S64(a)N+N66(a)T, V68(a)S, V68(a)T, R67(a)N+T69(a)S, R67(a)N, T69(a)N+M71(a)S, T69(a)N+M71(a)T, G72(a)N+F74(a)S, G72(a)N+F74(a)T, G73(a)N+K75(a)S, G73(a)N+K75(a)T, K75(a)N+E77(a)T, T86(a)N+Y88(a)S, T86(a)N+Y88(a)T, Y89(a)N+K91(a)S, Y89(a)N+K91(a)T, H90(a)N, and H90(a)N+S92(a)T, (having more than 50% side chain accessibility), and still more preferably from the group consiting of E9(a)N, F17(a)N, R67(a)N, and H90(a)N. The FSH-β part of this conjugate may be hFSH-β or any of the modified FSH-β polypeptides described herein.

Alternatively or additionally, the conjugate according to this embodiment comprises a modified FSH-β having an amino acid residue which differs from that of hFSH-β in the introduction of at least one N-glycosylation site by a mutation selected from the group consisting of S2(b)N+E4(b)S, S2(b)N+E4(b)T, E4(b)N+T6(b)S, E4(b)N, L5(b)N+N7(b)S,

L5(b)N+L7(b)T, T6(b)N+I8(b)S, T6(b)N+I8(b)T, I8(b)N+I10(b)S, I8(b)N+I10(b)T, T9(b)N+A11(b)S, T9(b)N+A11(b)T, K14(b)N+E16(b)S, K14(b)N+E16(b)T, F19(b)N+I21(b)S, F19(b)N+I21(b)T, I21(b)N+I23(b)S, I21(b)N+I23(b)T, S22(b)N+N24(b)S, S22(b)N+N24(b)T, Y31(b)N+Y33(b)S, Y31(b)N+Y33(b)T, Y33(b)N+R35(b)S, Y33(b)N+R35(b)T, R35(b)N+L37(b)S, R35(b)N+L37(b)T, D36(b)N+V38(b)S, D36(b)N+V38(b)T, L37(b)N+Y39(b)S, L37(b)N+Y39(b)T, K40(b)N+P42(b)S, K40(b)N+P42(b)T, A43(b)N+P45(b)S, A43(b)N+P45(b)T, P45(b)N+I47(b)S, P45(b)N+I47(b)T, K46(b)N+Q48(b)S, K46(b)N+Q48(b)T, I47(b)N+K49(b)S, I47(b)N+K49(b)T, K54(b)N+L56(b)S, K54(b)N+L56(b)T, E55(b)N+V57(b)S, E55(b)N+V57(b)T, L56(b)N+Y58(b)S, L56(b)N+Y58(b)T, V57(b)N+E59(b)S, V57(b)N+E59(b)T, Y58(b)N+T60(b)S, Y58(b)N, E59(b)N+V61(b)S, E59(b)N+V61(b)T, T60(b)N+R62(b)S, T60(b)N+R62(b)T, R62(b)N+P64(b)S, R62(b)N+P64(b)T, G65(b)N+A67(b)S, G65(b)N+A67(b)T, A67(b)N+H69(b)S, A67(b)N+H69(b)T, H68(b)N+A70(b)S, H68(b)N+A70(b)T, H69(b)N+D71(b)S, H69(b)N+D71(b)T, D71(b)N+L73(b)S, D71(b)N+L73(b)T, L73(b)N+T75(b)S, L73(b)N, T75(b)N+P77(b)S, T75(b)N+P77(b)T, H83(b)N+G85(b)S, H83(b)N+G85(b)T, K86(b)N+D88(b)S, K86(b)N+D88(b)T, D88(b)N+D90(b)S, D88(b)N+D90(b)T, S89(b)N, S89(b)N+S91(b)T, D90(b)N+T92(b)S, D90(b)N, S91(b)N+D93(b)S, S91(b)N+D93(b)T, D93(b)N+T96(b)S, D93(b)N, T95(b)N+R97(b)S, T95(b)N+R97(b)T, V96(b)N+G98(b)S, V96(b)N+G98(b)T, R97(b)N+L99(b)S, R97(b)N+L99(b)T, L99(b)N+P101(b)S, L99(b)N+P101(b)T, Y103(b)N, Y103(b)N+S105(b)T, S105(b)N+G107(b)S, S105(b)N+G107(b)T, F106(b)N+E108(b)S, F106(b)N+E108(b)T, G107(b)N+M109(b)S, G107(b)N+M109(b)T, E108(b)N+K110(b)S, E108(b)N+K110(b)T, M109(b)N+E111(b)S, and M109(b)N+E111(b)T, more preferably from the group consisting of E4(b)N, Y58(b)N, L73(b)N, S89(b)N, D90(b)N, D93(b)N, and Y103(b)N, even more preferably from the group consisting of F19(b)N+I21(b)S, F19(b)N+I21(b)T, Y33(b)N+R35(b)S, Y33(b)N+R35(b)T, A43(b)N+P45(b)S, A43(b)N+P45(b)T, P45(b)N+I47(b)S, P45(b)N+I47(b)T, K46(b)N+Q48(b)S, K46(b)N+Q48(b)T, I47(b)N+K49(b)S, I47(b)N+K49(b)T, K54(b)N+L56(b)S, K54(b)N+L56(b)T, E55(b)N+V57(b)S, E55(b)N+V57(b)T, V57(b)N+E59(b)S, V57(b)N+E59(b)T, Y58(b)N+T60(b)S, Y58(b)N, E59(b)N+V61(b)S, E59(b)N+V61(b)T, R62(b)N+P64(b)S, R62(b)N+P64(b)T, G65(b)N+A67(b)S, G65(b)N+A67(b)T, A67(b)N+H69(b)S, A67(b)N+H69(b)T, H68(b)N+A70(b)S, H68(b)N+A70(b)T, H69(b)N+D71(b)S, H69(b)N+D71(b)T, D71(b)N+L73(b)S, D71(b)N+L73(b)T, L73(b)N+T75(b)S, L73(b)N, T75(b)N+P77(b)S, T75(b)N+P77(b)T, H83(b)N+G85(b)S, H83(b)N+G85(b)T, K86(b)N+D88(b)S, K86(b)N+D88(b)T, D88(b)N+D90(b)S, D88(b)N+D90(b)T, S89(b)N, S89(b)N+S91(b)T, D90(b)N+T92(b)S, D90(b)N, S91(b)N+D93(b)S, S91(b)N+D93(b)T, T95(b)N+R97(b)S, T95(b)N+R97(b)T, R97(b)N+L99(b)S, R97(b)N+L99(b)T, L99(b)N+P101(b)S, L99(b)N+P101(b)T, Y103(b)N, Y103(b)N+S105(b)T, S105(b)N+G107(b)S, S105(b)N+G107(b)T, F106(b)N+E108(b)S, F106(b)N+E108(b)T, G107(b)N+M109(b)S, G107(b)N+M109(b)T, E108(b)N+K110(b)S, E108(b)N+K110(b)T, M109(b)N+E111(b)S, and M109(b)N+E111(b)T (having more than 50% side chain accessibility), and even more preferably from the group consisting of Y58(b)N, L73(b)N, S89(b)N, D90(b)N, and Y103(b)N. The FSH-α part of this conjugate may be hFSH-α or any of the modified FSH-α polypeptides described herein.

The FSH-α and/or FSH-β polypeptide may further differ from hFSH-α and/or hFSH-β in at least one removed, naturally occurring N-glycosylation site. In particular FSH-α may comprise a substitution of N78(a) and/or T80(a) by any other amino acid residue and/or FSH-β a

substitution of N7(b), T9(b), N24(b) and/or T26(b) by any other amino acid residue. Preferably, the N residue is substituted by Q or D, and the T residue by A or G.

Furthermore, FSH-α of the conjugate according to this embodiment (having at least one of the above mentioned N-glycosylation site modifications) may differ from hFSH-α in the removal of at least one lysine residue selected from the group consisting of K44(a), K45(a), K51(a), K63(a), K75(a), and K91(a), in particular at least one amino acid residue selected from of the group consisting of K44(a), K45(a), K63(a), K75(a), and K91(a) (these residues having more than 25% of their side chain exposed to the surface), and preferably from the group consisting of K45(a), K63(a), K75(a), and K91(a) (these residues having more than 50% of their side chain exposed to the surface).

An alternative embodiment of this aspect of the invention is one in which at least one of said FSH-α and FSH-β subunits comprises at least one introduced N- or O-glycosylation site at the N-terminal thereof, and wherein the at least one introduced glycosylation site is glycosylated. In this case, the respective subunits may comprise one or more of the modifications disclosed elsewhere herein, or one or both of the subunits may be the respective wildtype subunits, but having the at least one introduced terminal glycosylation site. Thus, the polypeptide conjugate may be one in which the FSH-α subunit comprises hFSH-α having the sequence shown in SEQ ID NO 2, and/or in which the FSH-β subunit comprises hFSH-β having the sequence shown in SEQ ID NO 4. In a particular embodiment, both of the subunits correspond to the respective wildtype hFSH subunits, although with either the α or β subunit, or both, having an introduced N-terminal glycosylation site.

The introduced glycosylation site may be of the type described elsewhere herein; see the discussion of glycosylation under the general discussion of attachment groups above. A non-limiting example of a suitable glycosylation site for introduction at the N-terminal is the sequence Ala-Asn-Ile-Thr-Val-Asn-Ile-Thr-Val, e.g. for insertion upstream of a mature FSH-α sequence.

It will be understood that in order to prepare a conjugate according to this aspect the polypeptide i) must be expressed in a glycosylating host cell capable of attaching oligosaccharide moieties at the glycosylation site(s) or alternatively subjected to *in vitro* glycosylation. Examples of glycosylating host cells are given in the section further below entitled "Coupling to an oligosaccharide moiety".

In addition to a carbohydrate molecule, the conjugate according to the aspect of the invention described in the present section may contain additional non-polypeptide moieties different from O-linked or N-linked carbohydrate moieties, in particular a polymer molecule as described herein conjugated to one or more attachment groups present in the polypeptide part of the conjugate. This is particularly relevant when a lysine residue (or any other amino acid residue comprising an attachment group for the non-polypeptide moiety in question) has been introduced and/or removed.

It will be understood that any of the amino acid changes, in particular substitutions, specified in this section can be combined with any of the amino acid changes, in particular substitutions, specified in the other sections herein disclosing specific amino acid changes.

Non-polypeptide moiety of the conjugate of the invention

As indicated further above the non-polypeptide moiety of the conjugate of the invention is preferably selected from the group consisting of a polymer molecule, a lipophilic compound, an oligosaccharide moiety (by way of *in vivo* glycosylation) and an organic derivatizing agent. All of these agents may confer desirable properties to the polypeptide part of the conjugate, in particular an increased functional *in vivo* half-life and/or an increased serum half-life. The polypeptide part of the conjugate is normally conjugated to only one type of non-polypeptide moiety, but may also be conjugated to two or more different types of non-polypeptide moieties, e.g. to a polymer molecule and an oligosaccharide moiety, to a lipophilic group and an oligosaccharide moiety, to an organic derivatizing agent and an oligosaccharide moiety, to a lipophilic group and a polymer molecule, etc. The conjugation to two or more different non-polypeptide moieties may be done simultaneous or sequentially.

Polypeptide of the invention

In a further aspect the invention relates to a modified FSH-α or a modified FSH-β polypeptide constituting part of a conjugate of the invention. The modified FSH-α and FSH-β is preferably glycosylated and thus further comprises N-linked and/or O-linked oligosaccharide moieties. Specific modified FSH-α and FSH-β polypeptides of the invention are those described in the section entitled "Conjugate of the invention".

Methods of preparing a conjugate of the invention

In the following sections "Conjugation to a lipophilic compound", "Conjugation to a polymer molecule", "Conjugation to an oligosaccharide moiety" and "Conjugation to an organic derivatizing agent", conjugation to specific types of non-polypeptide moieties is described.

Conjugation to a lipophilic compound

The polypeptide and the lipophilic compound may be conjugated to each other, either directly or by use of a linker. The lipophilic compound may be a natural compound such as a saturated or unsaturated fatty acid, a fatty acid diketone, a terpene, a prostaglandin, a vitamin, a carotenoid or steroid, or a synthetic compound such as a carbon acid, an alcohol, an amine and sulphonic acid with one or more alkyl, aryl, alkenyl or other multiple unsaturated compounds. The conjugation between the polypeptide and the lipophilic compound, optionally through a linker, may be done according to methods known in the art, e.g. as described by Bodanszky in Peptide Synthesis, John Wiley, New York, 1976 and in WO 96/12505.

Conjugation to a polymer molecule

The polymer molecule to be coupled to the polypeptide may be any suitable polymer molecule, such as a natural or synthetic homo-polymer or hetero-polymer, typically with a molecular weight in the range of 300-50,000 Da, such as 300-20,000 Da, more preferably in the range of 500-10,000 Da, even more preferably in the range of 500-5000 Da. Examples of homo-polymers include a polyol (i.e. poly-OH), a polyamine (i.e. poly-NH_2) and a polycarboxylic acid (i.e. poly-COOH). A hetero-polymer is a polymer which comprises different coupling groups, such as a hydroxyl group and an amine group.

Examples of suitable polymer molecules include polymer molecules selected from the group consisting of polyalkylene oxide (PAO), including polyalkylene glycol (PAG), such as polyethylene glycol (PEG) and polypropylene glycol (PPG), branched PEGs, poly-vinyl alcohol (PVA), poly-carboxylate, poly-(vinylpyrolidone), polyethylene-co-maleic acid anhydride, polystyrene-co-maleic acid anhydride, dextran, including carboxymethyl-dextran, or any other biopolymer suitable for reducing immunogenicity and/or increasing functional *in vivo* half-life and/or serum half-life. Another example of a polymer molecule is human albumin or

another abundant plasma protein. Generally, polyalkylene glycol-derived polymers are biocompatible, non-toxic, non-antigenic, non-immunogenic, have various water solubility properties, and are easily excreted from living organisms.

PEG is the preferred polymer molecule, since it has only few reactive groups capable of cross-linking compared, e.g., to polysaccharides such as dextran, and the like. In particular, monofunctional PEG, e.g. methoxypolyethylene glycol (mPEG), is of interest since its coupling chemistry is relatively simple (only one reactive group is available for conjugating with attachment groups on the polypeptide). Consequently, the risk of cross-linking is eliminated, the resulting polypeptide conjugates are more homogeneous and the reaction of the polymer molecules with the polypeptide is easier to control.

To effect covalent attachment of the polymer molecule(s) to the polypeptide, the hydroxyl end groups of the polymer molecule must be provided in activated form, i.e. with reactive functional groups. Suitable activated polymer molecules are commercially available, e.g. from Shearwater Polymers, Inc., Huntsville, AL, USA. Alternatively, the polymer molecules can be activated by conventional methods known in the art, e.g. as disclosed in WO 90/13540. Specific examples of activated linear or branched polymer molecules for use in the present invention are described in the Shearwater Polymers, Inc. 1997 and 2000 Catalogs (Functionalized Biocompatible Polymers for Research and pharmaceuticals, Polyethylene Glycol and Derivatives, incorporated herein by reference). Specific examples of activated PEG polymers include the following linear PEGs: NHS-PEG (e.g. SPA-PEG, SSPA-PEG, SBA-PEG, SS-PEG, SSA-PEG, SC-PEG, SG-PEG, and SCM-PEG), and NOR-PEG), BTC-PEG, EPOX-PEG, NCO-PEG, NPC-PEG, CDI-PEG, ALD-PEG, TRES-PEG, VS-PEG, IODO-PEG, and MAL-PEG, and branched PEGs such as PEG2-NHS and those disclosed in US 5,932,462 and US 5,643,575, both of which are incorporated herein by reference. Furthermore, the following publications, incorporated herein by reference, disclose useful polymer molecules and/or PEGylation chemistries: US 5,824,778, US 5,476,653, WO 97/32607, EP 229,108, EP 402,378, US 4,902,502, US 5,281,698, US 5,122,614, US 5,219,564, WO 92/16555, WO 94/04193, WO 94/14758, WO 94/17039, WO 94/18247, WO 94/28024, WO 95/00162, WO 95/11924, WO95/13090, WO 95/33490, WO 96/00080, WO 97/18832, WO 98/41562, WO 98/48837, WO 99/32134, WO 99/32139, WO 99/32140, WO 96/40791, WO 98/32466, WO 95/06058, EP 439 508, WO 97/03106, WO 96/21469, WO 95/13312, EP 921 131, US 5,736,625, WO 98/05363, EP 809 996, US 5,629,384, WO 96/41813, WO 96/07670, US 5,473,034, US 5,516,673, EP 605 963, US 5,382,657, EP 510 356, EP 400 472, EP 183 503 and EP 154 316.

The conjugation of the polypeptide and the activated polymer molecules is conducted by use of any conventional method, e.g. as described in the following references (which also describe suitable methods for activation of polymer molecules): R.F. Taylor, (1991), "Protein immobilisation. Fundamental and applications", Marcel Dekker, N.Y.; S.S. Wong, (1992), "Chemistry of Protein Conjugation and Crosslinking", CRC Press, Boca Raton; G.T. Hermanson et al., (1993), "Immobilized Affinity Ligand Techniques", Academic Press, N.Y.). The skilled person will be aware that the activation method and/or conjugation chemistry to be used depends on the attachment group(s) of the polypeptide (examples of which are given further above), as well as the functional groups of the polymer (e.g. being amine, hydroxyl, carboxyl, aldehyde, sulfydryl, succinimidyl, maleimide, vinysulfone or haloacetate). The PEGylation may be directed towards conjugation to all available attachment groups on the polypeptide (i.e. such attachment groups that are exposed at the surface of the polypeptide) or may be directed towards one or more specific attachment groups, e.g. the N-terminal amino

group (US 5,985,265). Furthermore, the conjugation may be achieved in one step or in a stepwise manner (e.g. as described in WO 99/55377).

It will be understood that the PEGylation is designed so as to produce the optimal molecule with respect to the number of PEG molecules attached, the size and form of such molecules (e.g. whether they are linear or branched), and the attachment site(s) in the polypeptide. The molecular weight of the polymer to be used may e.g. be chosen on the basis of the desired effect to be achieved. For instance, in order to obtain reduced renal clearance (and thus increased half-life of the conjugate), the molecular weight of the conjugate is important. Accordingly, for this purpose the PEGylation is designed so as to achieve a sufficiently high molecular weight of the conjugate, e.g. a molecular weight of at least about 67 kDa in many cases. As indicated above, in other cases is may however be desirable to have a molecular weight that is somewhat increased, but which still is below about 67 kDa. In such cases, PEGylation may be performed so as to produce conjugates having one or more relatively small PEG polymers, for example one, two or three PEG polymers each having a molecular weight of e.g. up to about 5000 Da.

In connection with conjugation to only a single attachment group on the protein (as described in US 5,985,265), it may be advantageous that the polymer molecule, which may be linear or branched, has a high molecular weight, e.g. about 20 kDa.

In a specific embodiment, the polypeptide conjugate of the invention is one which comprises a single PEG molecule attached to the N-terminal of the polypeptide and no other PEG molecules, in particular a linear or branched PEG molecule with a molecular weight of at least about 20 kDa. The polypeptide according to this embodiment may further comprise one or more oligosaccharide moieties attached to an N-linked or O-linked glycosylation site of the polypeptide or carbohydrate moieties attached by *in vitro* glycosylation.

In another specific embodiment, the polypeptide conjugate of the invention comprises a PEG molecule attached to each of the lysine residues in the polypeptide available for PEGylation, in particular a linear or branched PEG molecule, e.g. with a molecular weight of about 5 kDa.

In yet another embodiment, the polypeptide conjugate of the invention comprises a PEG molecule attached to each of the lysine residues in the polypeptide available for PEGylation, and in addition to the N-terminal amino acid residue of the polypeptide.

Normally, the polymer conjugation is performed under conditions aiming at reacting all available polymer attachment groups with polymer molecules. Typically, the molar ratio of activated polymer molecules to polypeptide is up to 500-1, such as 200-1, preferably 100-1, such as 50-1 or 25-1 in order to obtain optimal reaction. Furthermore, the polymer modification, such as PEGylation, is conveniently carried out at at a pH in the range of 7-10, such as in the range of 8-10, in particular in the range of 8-9.

It is also contemplated according to the invention to couple the polymer molecules to the polypeptide through a linker. Suitable linkers are well known to the skilled person. A preferred example is cyanuric chloride (Abuchowski *et al.*, (1977), J. Biol. Chem., 252, 3578-3581; US 4,179,337; Shafer *et al.*, (1986), J. Polym. Sci. Polym. Chem. Ed., 24, 375-378.

Subsequent to the conjugation residual activated polymer molecules are blocked according to methods known in the art, e.g. by addition of primary amine to the reaction mixture, and the resulting inactivated polymer molecules removed by a suitable method.

Covalent *in vitro* coupling of carbohydrate moieties glycosides (such as dextran) to amino acid residues of the polypeptide may also be used, e.g. as described in WO 87/05330 and in Aplin et al., CRC Crit Rev. Biochem., pp. 259-306, 1981. The *in vitro* coupling of carbohydrate moieties or PEG to protein- and peptide-bound Gln-residues can be carried out by transglutaminases (TGases). Transglutaminases catalyse the transfer of donor amine-groups to protein- and peptide-bound Gln residues in a so-called cross-linking reaction. The donor-amine groups can be protein- or peptide-bound e.g. as the ε-amino-group in Lys residues or can be part of a small or large organic molecule. An example of a small organic molecule functioning as amino-donor in TGase-catalysed cross-linking is putrescine (1,4-diaminobutane). An example of a larger organic molecule functioning as amino-donor in TGase-catalysed cross-linking is an amine-containing PEG (Sato et al., Biochemistry 35, 13072-13080).

TGases, in general, are highly specific enzymes, and not every Gln residue exposed on the surface of a protein is accessible to TGase-catalysed cross-linking to amino-containing substances. On the contrary, only a few Gln residues function naturally as TGase substrates but the exact parameters governing which Gln residues are good TGase substrates remain unknown. Thus, in order to render a protein susceptible to TGase-catalysed cross-linking reactions it is often a prerequisite at convenient positions to add stretches of amino acid sequence known to function very well as TGase substrates. Several amino acid sequences are known to be or to contain excellent natural TGase substrates e.g. substance P, elafin, fibrinogen, fibronectin, α_2-plasmin inhibitor, α-caseins, and β-caseins.

Coupling to an oligosaccharide moiety

The conjugation to an oligosaccharide moiety takes place by *in vivo* glycosylation effected by a glycosylating, eucaryotic expression host. The expression host cell may be selected from fungal (filamentous fungal or yeast), insect or animal cells or from transgenic plant cells. In one embodiment the host cell is a mammalian cell, such as a CHO cell, BHK or HEK, e.g. HEK 293, cell, or an insect cell, such as an SF9 cell, or a yeast cell, e.g. *S. cerevisiae* or *Pichia pastoris*, or any of the host cells mentioned hereinafter.

Coupling to an organic derivatizing agent

Covalent modification of the polypeptide exhibiting FSH activity may be performed by reacting one or more (attachment groups of the polypeptide with an organic derivatizing agent. Suitable derivatizing agents and methods are well known in the art. For example, cysteinyl residues most commonly are reacted with α-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, α-bromo-β-(4-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole. Histidyl residues are derivatized by reaction with diethylpyrocarbonateat, pH 5.5-7.0, because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide is also useful. The reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0. Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for deri-

vatizing α-amino-containing residues include imidoesters such as methyl picolinimidate, pyridoxal phosphate, pyridoxal, chloroborohydride, trinitrobenzenesulfonic acid, O-methylisourea, 2,4-pentanedione and transaminase-catalyzed reaction with glyoxylate. Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed under alkaline conditions because of the high pKa of the guanidine functional group.

Furthermore, these reagents may react with the groups of lysine as well as the arginine guanidino group. Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimides (R-N=C=N-R'), where R and R' are different alkyl groups, such as 1-cyclohexyl-3-(2-morpholinyl-4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Blocking of a functional site

It has been reported that excessive polymer conjugation can lead to a loss of activity of the polypeptide to which the polymer is conjugated. This problem can be eliminated, e.g., by removal of attachment groups located at the functional site or by blocking the functional site prior to conjugation. The latter strategy constitutes a further embodiment of the invention (the first strategy being exemplified further above, e.g. by removal of lysine residues which may be located close to the functional site). More specifically, according to the second strategy the conjugation between the polypeptide and the non-polypeptide moiety ii) is conducted under conditions where the functional site of the polypeptide i) is blocked by a helper molecule capable of binding to the functional site of the polypeptide i).

Preferably, the helper molecule is one which specifically recognizes a functional site of the polypeptide, such as a receptor, in particular the FSH receptor or a part of the FSH receptor. Alternatively, the helper molecule may be an antibody, in particular a monoclonal antibody recognizing the polypeptide exhibiting FSH activity. In particular, the helper molecule may be a neutralizing monoclonal antibody.

The polypeptide is allowed to interact with the helper molecule before effecting conjugation. This ensures that the functional site of the polypeptide is shielded or protected and consequently unavailable for derivatization by the non-polypeptide moiety such as a polymer. Following its elution from the helper molecule, the conjugate between the non-polypeptide moiety and the polypeptide can be recovered with at least a partially preserved functional site.

The subsequent conjugation of the polypeptide having a blocked functional site to a polymer, a lipophilic compound, an oligosaccharide moiety, an organic derivatizing agent or any other compound is conducted in the normal way, e.g. as described in the sections above entitled "Conjugation to ....".

Irrespective of the nature of the helper molecule to be used to shield the functional site of the polypeptide from conjugation, it is desirable that the helper molecule is free of or comprises only a few attachment groups for the non-polypeptide moiety of choice in any parts of the molecule where the conjugation to such groups will hamper the desorption of the conjugated polypeptide from the helper molecule. Hereby, selective conjugation to attachment groups present in non-shielded parts of the polypeptide can be obtained and it is possible to reuse the helper molecule for repeated cycles of conjugation. For instance, if the non-polypeptide moiety is a polymer molecule such as PEG which has the epsilon amino group of a lysine or N-

terminal amino acid residue as an attachment group, it is desirable that the helper molecule is substantially free of conjugatable epsilon amino groups, preferably free of any epsilon amino groups. Accordingly, in a preferred embodiment the helper molecule is a protein or peptide capable of binding to the functional site of the polypeptide, which protein or peptide is free of any conjugatable attachment groups for the non-polypeptide moiety of choice.

In a further embodiment the helper molecule is first covalently linked to a solid phase such as column packing materials, for instance Sephadex or agarose beads, or a surface, e.g. a reaction vessel. Subsequently, the polypeptide is loaded onto the column material carrying the helper molecule and conjugation carried out according to methods known in the art, e.g. as described in the sections above entitled "Conjugation to ....". This procedure allows the polypeptide conjugate to be separated from the helper molecule by elution. The polypeptide conjugate is eluated by conventional techniques under physico-chemical conditions that do not lead to a substantive degradation of the polypeptide conjugate. The fluid phase containing the polypeptide conjugate is separated from the solid phase to which the helper molecule remains covalently linked. The separation can be achieved in other ways: For instance, the helper molecule may be derivatised with a second molecule (e.g. biotin) that can be recognized by a specific binder (e.g. streptavidin). The specific binder may be linked to a solid phase thereby allowing the separation of the polypeptide conjugate from the helper molecule-second molecule complex through passage over a second helper-solid phase column which will retain, upon subsequent elution, the helper molecule-second molecule complex, but not the polypeptide conjugate. The polypeptide conjugate may be released from the helper molecule in any appropriate fashion. Deprotection may be achieved by providing conditions in which the helper molecule dissociates from the functional site of the FSH to which it is bound. For instance, a complex between an antibody to which a polymer is conjugated and an anti-idiotypic antibody can be dissociated by adjusting the pH to an acid or alkaline pH.

Conjugation of a tagged polypeptide

In an alternative embodiment the polypeptide i) is expressed as a fusion protein with a tag, i.e. an amino acid sequence or peptide stretch made up of typically 1-30, such as 1-20 amino acid residues. Besides allowing for fast and easy purification, the tag is a convenient tool for achieving conjugation between the tagged polypeptide i) and the non-polypeptide moiety ii). In particular, the tag may be used for achieving conjugation in microtiter plates or other carriers, such as paramagnetic beads, to which the tagged polypeptide can be immobilised via the tag. The conjugation to the tagged polypeptide i) in, e.g., microtiter plates has the advantage that the tagged polypeptide can be immobilised in the microtiter plates directly from the culture broth (in principle without any purification) and subjected to conjugation. Thereby, the total number of process steps (from expression to conjugation) can be reduced. Furthermore, the tag may function as a spacer molecule ensuring an improved accessibility to the immobilised polypeptide to be conjugated. The conjugation using a tagged polypeptide i) may be to any of the non-polypeptide moieties disclosed herein, e.g. to a polymer molecule such as PEG.

The identity of the specific tag to be used is not critical as long as the tag is capable of being expressed with the polypeptide i) and is capable of being immobilised on a suitable surface or carrier material. A number of suitable tags are commercially available, e.g. from Unizyme Laboratories, Denmark. For instance, the tag may consist of any of the following sequences:
His-His-His-His-His-His
Met-Lys-His-His-His-His-His-His
Met-Lys-His-His-Ala-His-His-Gln-His-His
Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln

Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-Gln

or any of the following:
EQKLI SEEDL (a C-terminal tag described in Mol. Cell. Biol. 5:3610-16, 1985)
DYKDDDDK (a C- or N-terminal tag)
YPYDVPDYA

Antibodies against the above tags are commercially available, e.g. from ADI, Aves Lab and Research Diagnostics.

The subsequent cleavage of the tag from the polypeptide i) may be achieved by use of commercially available enzymes.

Methods for preparing a polypeptide of the invention or the polypeptide i) of the conjugate of the invention

The polypeptide of the present invention or the polypeptide part of a conjugate of the invention, optionally in glycosylated form, may be produced by any suitable method known in the art. Such methods include constructing a nucleotide sequence encoding the polypeptide and expressing the sequence in a suitable transformed or transfected host. Polypeptides of the invention may also be produced, albeit less efficiently, by chemical synthesis or a combination of chemical synthesis and recombinant DNA technology.

FSH-α and FSH-β may be expressed separately and subquently allowed to dimerize. However, it is preferred that FSH-α and FSH-β are expressed by the same host cell and dimerized *in vivo* prior to purification and any conjugation to a non-polypeptide moiety. Co-expression of FSH-α and FSH-β in CHO cells is described by Keene et al., J Biol Chem 1989 25; 264(9): 4769-75. Alternatively, the polypeptide i) may be expressed as a single-chain polypeptide wherein the nucleotide sequences encoding FSH-α and FSH-β are fused, directly or using a suitable linker, and expressed as a single-chain polypeptide using a similar approach to that described in US 5,883,073.

The nucleotide sequence encoding FSH-α or FSH-β modified according to the invention may be constructed by isolating or synthesizing a nucleotide sequence encoding the parent FSH subunit, such as hFSH-α or hFSH-β with the amino acid sequence shown in SEQ ID NO 2 or 4, respectively, or the precursor form thereof (shown in SEQ ID NO 1 and 3, respectively) and then changing the nucleotide sequence so as to effect introduction (i.e. insertion or substitution) or deletion (i.e. removal or substitution) of the relevant amino acid residue(s). The nucleotide sequence is conveniently modified by site-directed mutagenesis in accordance with conventional methods. Alternatively, the nucleotide sequence may be prepared by chemical synthesis, e.g. by using an oligonucleotide synthesizer, wherein oligonucleotides are designed based on the amino acid sequence of the desired polypeptide, and preferably selecting those codons that are favored in the host cell in which the recombinant polypeptide will be produced. For example, several small oligonucleotides coding for portions of the desired polypeptide may be synthesized and assembled by PCR, ligation or ligation chain reaction (LCR) (Barany, PNAS 88:189-193, 1991). The individual oligonucleotides typically contain 5' or 3' overhangs for complementary assembly.

Once assembled (by synthesis, site-directed mutagenesis or another method), the nucleotide sequence encoding the polypeptide is inserted into a recombinant vector and operably linked to control sequences necessary for expression of the FSH in the desired transformed host cell.

It should of course be understood that not all vectors and expression control sequences function equally well to express the nucleotide sequence encoding a polypeptide described herein. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation. For example, in selecting a vector, the host must be considered because the vector must replicate in it or be able to integrate into the chromosome. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleotide sequence encoding the polypeptide, particularly as regards potential secondary structures. Hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleotide sequence, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification of the products coded for by the nucleotide sequence.

The recombinant vector may be an autonomously replicating vector, i.e. a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector is one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the nucleotide sequence encoding the polypeptide of the invention is operably linked to additional segments required for transcription of the nucleotide sequence. The vector is typically derived from plasmid or viral DNA. A number of suitable expression vectors for expression in the host cells mentioned herein are commercially available or described in the literature. Useful expression vectors for eukaryotic hosts include, for example, vectors comprising expression control sequences from SV40, bovine papilloma virus, adenovirus and cytomegalovirus. Specific vectors are, e.g., pCDNA3.1(+)\Hyg (Invitrogen, Carlsbad, CA, USA) and pCI-neo (Stratagene, La Jolla, CA, USA). Useful expression vectors for yeast cells include the 2μ plasmid and derivatives thereof, the POT1 vector (US 4,931,373), the pJSO37 vector described in Okkels, Ann. New York Acad. Sci. 782, 202-207, 1996, and pPICZ A, B or C (Invitrogen). Useful vectors for insect cells include pVL941, pBG311 (Cate *et al.*, "Isolation of the Bovine and Human Genes for Mullerian Inhibiting Substance And Expression of the Human Gene In Animal Cells", Cell, 45, pp. 685-98 (1986), pBluebac 4.5 and pMelbac (both available from Invitrogen). Useful expression vectors for bacterial hosts include known bacterial plasmids, such as plasmids from *E. coli*, including pBR322, pET3a and pET12a (both from Novagen Inc., WI, USA), wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g. , NM989, and other DNA phages, such as M13 and filamentous single stranded DNA phages.

Other vectors for use in this invention include those that allow the nucleotide sequence encoding the polypeptide to be amplified in copy number. Such amplifiable vectors are well known in the art. They include, for example, vectors able to be amplified by DHFR amplification (see, e.g., Kaufman, U.S. Pat. No. 4,470,461, Kaufman and Sharp, "Construction Of A Modular Dihydrafolate Reductase cDNA Gene: Analysis Of Signals Utilized For Efficient Expression", Mol. Cell. Biol., 2, pp. 1304-19 (1982)) and glutamine synthetase ("GS") amplification (see, e.g., US 5,122,464 and EP 338,841).

In a preferred embodiment a pair of expression vectors are used for expressing the polypeptide i) of the invention or constituting part of a conjugate of the invention. Each of the vectors of said pair is capable of transfecting an eukaryotic cell as described herein, and the vectors comprise nucleotide sequences encoding, respectively, a modified FSH-α as described herein and a wildtype FSH-β subunit, a modified FSH-β as described herein and a wildtype FSH-α subunit, or a modified FSH-α and a modified FSH-β as described herein. The use of a pair of vectors is, e.g., described in EP 211,894.

The recombinant vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV40 origin of replication. When the host cell is a yeast cell, suitable sequences enabling the vector to replicate are the yeast plasmid 2μ replication genes REP 1-3 and origin of replication.

The vector may also comprise a selectable marker, e.g. a gene whose product complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the Schizosaccharomyces pombe TPI gene (described by P.R. Russell, Gene 40, 1985, pp. 125-130), or one which confers resistance to a drug, e.g. ampicillin, kanamycin, tetracyclin, chloramphenicol, neomycin, hygromycin or methotrexate. For *Saccharomyces cerevisiae*, selectable markers include *ura3 and leu2*. For filamentous fungi, selectable markers include *amdS*, *pyrG*, *arcB*, *niaD* and *sC*.

The term "control sequences" is defined herein to include all components which are necessary or advantageous for the expression of the polypeptide of the invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader sequence, polyadenylation sequence, propeptide sequence, promoter, enhancer or upstream activating sequence, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter.

A wide variety of expression control sequences may be used in the present invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors as well as any sequence known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

Examples of suitable control sequences for directing transcription in mammalian cells include the early and late promoters of SV40 and adenovirus, e.g. the adenovirus 2 major late promoter, the MT-1 (metallothionein gene) promoter, the human cytomegalovirus immediate-early gene promoter (CMV), the human elongation factor 1α (EF-1α) promoter, the *Drosophila* minimal heat shock protein 70 promoter, the Rous Sarcoma Virus (RSV) promoter, the human ubiquitin C (UbC) promoter, the human growth hormone terminator, SV40 or adenovirus Elb region polyadenylation signals and the Kozak consensus sequence (Kozak, M. *J Mol Biol* 1987 Aug 20;196(4):947-50).

In order to improve expression in mammalian cells a synthetic intron may be inserted in the 5' untranslated region of the nucleotide sequence encoding the polypeptide. An example of a synthetic intron is the synthetic intron from the plasmid pCI-Neo (available from Promega Corporation, WI, USA).

Examples of suitable control sequences for directing transcription in insect cells include the polyhedrin promoter, the P10 promoter, the *Autographa californica* polyhedrosis virus basic protein promoter, the baculovirus immediate early gene 1 promoter and the baculovirus 39K delayed-early gene promoter, and the SV40 polyadenylation sequence. Examples of suitable control sequences for use in yeast host cells include the promoters of the yeast α-mating system, the yeast triose phosphate isomerase (TPI) promoter, promoters from yeast glycolytic genes or alcohol dehydrogenase genes, the ADH2-4c promoter, and the inducible GAL promoter. Examples of suitable control sequences for use in filamentous fungal host cells include the ADH3 promoter and terminator, a promoter derived from the genes encoding *Aspergillus oryzae* TAKA amylase triose phosphate isomerase or alkaline protease, an *A. niger* α-amylase, *A. niger* or *A. nidulans* glucoamylase, *A. nidulans* acetamidase, *Rhizomucor miehei* aspartic proteinase or lipase, the TPI1 terminator and the ADH3 terminator. Examples of suitable control sequences for use in bacterial host cells include promoters of the *lac* system, the *trp* system, the *TAC* or *TRC* system, and the major promoter regions of phage lambda.

The presence or absence of a signal peptide will, e.g., depend on the expression host cell used for the production of the polypeptide to be expressed (whether it is an intracellular or extracellular polypeptide) and whether it is desirable to obtain secretion. For use in filamentous fungi, the signal peptide may conveniently be derived from a gene encoding an *Aspergillus* sp. amylase or glucoamylase, a gene encoding a *Rhizomucor miehei* lipase or protease or a *Humicola lanuginosa* lipase. The signal peptide is preferably derived from a gene encoding *A. oryzae* TAKA amylase, *A. niger* neutral α-amylase, *A. niger* acid-stable amylase, or *A. niger* glucoamylase. For use in insect cells, the signal peptide may conveniently be derived from an insect gene (cf. WO 90/05783), such as the *Lepidopteran manduca sexta* adipokinetic hormone precursor, (cf. US 5,023,328), the honeybee melittin (Invitrogen), ecdysteroid UDPglucosyltransferase (egt) (Murphy *et al.*, Protein Expression and Purification 4, 349-357 (1993) or human pancreatic lipase (hpl) (Methods in Enzymology 284, pp. 262-272, 1997). A preferred signal peptide for use in mammalian cells is that of hFSH or the murine Ig kappa light chain signal peptide (Coloma, M (1992) J. Imm. Methods 152:89-104). For use in yeast cells suitable signal peptides have been found to be the α-factor signal peptide from *S. cerevisiae* (cf. US 4,870,008), a modified carboxypeptidase signal peptide (cf. L.A. Valls *et al.*, Cell 48, 1987, pp. 887-897), the yeast BAR1 signal peptide (cf. WO 87/02670), the yeast aspartic protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani *et al.*, Yeast 6, 1990, pp. 127-137), and the synthetic leader sequence TA57 (WO98/32867). For use in *E. coli* cells a suitable signal peptide have been found to be the signal peptide *ompA* (EP581821).

The nucleotide sequence of the invention encoding a polypeptide exhibiting FSH activity, whether prepared by site-directed mutagenesis, synthesis, PCR or other methods, may optionally also include a nucleotide sequence that encodes a signal peptide. The signal peptide is present when the polypeptide is to be secreted from the cells in which it is expressed. Such signal peptide, if present, should be one recognized by the cell chosen for expression of the polypeptide. The signal peptide may be homologous (e.g. be that normally associated with a hFSH subunit) or heterologous (i.e. originating from another source than hFSH) to the polypeptide or may be homologous or heterologous to the host cell, i.e. be a signal peptide normally expressed from the host cell or one which is not normally expressed from the host cell. Accordingly, the signal peptide may be prokaryotic, e.g. derived from a bacterium such as *E. coli,* or eukaryotic, e.g. derived from a mammalian, or insect or yeast cell.

Any suitable host may be used to produce the polypeptide or polypeptide part of the conjugate of the invention, including bacteria, fungi (including yeasts), plant, insect, mammal, or other appropriate animal cells or cell lines, as well as transgenic animals or plants. Examples of bacterial host cells include gram-positive bacteria such as strains of *Bacillus*, e.g. *B. brevis* or *B. subtilis*, *Pseudomonas* or *Streptomyces*, or gram-negative bacteria, such as strains of *E. coli*. The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278). Examples of suitable filamentous fungal host cells include strains of *Aspergillus*, e.g. *A. oryzae*, *A. niger*, or *A. nidulans*, *Fusarium* or *Trichoderma*. Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and US 5,679,543. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156 and WO 96/00787. Examples of suitable yeast host cells include strains of *Saccharomyces*, e.g. *S. cerevisiae*, *Schizosaccharomyces*, *Klyveromyces*, *Pichia*, such as *P. pastoris* or *P. methanolica*, *Hansenula*, such as *H. Polymorpha* or *Yarrowia*. Yeast may be transformed using the procedures described by Becker and Guarente, *In* Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, *Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920: and as disclosed by Clontech Laboratories, Inc, Palo Alto, CA, USA (in the product protocol for the Yeastmaker™ Yeast Transformation System Kit). Examples of suitable insect host cells include a *Lepidoptora* cell line, such as *Spodoptera frugiperda* (Sf9 or Sf21) or *Trichoplusioa ni* cells (High Five) (US 5,077,214). Transformation of insect cells and production of heterologous polypeptides therein may be performed as described by Invitrogen. Examples of suitable mammalian host cells include Chinese hamster ovary (CHO) cell lines, (e.g. CHO-K1; ATCC CCL-61), Green Monkey cell lines (COS) (e.g. COS 1 (ATCC CRL-1650), COS 7 (ATCC CRL-1651)); mouse cells (e.g. NS/O), Baby Hamster Kidney (BHK) cell lines (e.g. ATCC CRL-1632 or ATCC CCL-10), and human cells (e.g. HEK 293 (ATCC CRL-1573)), as well as plant cells in tissue culture. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. Methods for introducing exogeneous DNA into mammalian host cells include calcium phosphate-mediated transfection, electroporation, DEAE-dextran mediated transfection, liposome-mediated transfection, viral vectors and the transfection method described by Life Technologies Ltd, Paisley, UK using Lipofectamin 2000. These methods are well known in the art and e.g. described by Ausbel *et al.* (eds.), 1996, Current Protocols in Molecular Biology, John Wiley & Sons, New York, USA. The cultivation of mammalian cells are conducted according to established methods, e.g. as disclosed in (Animal Cell Biotechnology, Methods and Protocols, Edited by Nigel Jenkins, 1999, Human Press Inc, Totowa, New Jersey, USA and Harrison MA and Rae IF, General Techniques of Cell Culture, Cambridge University Press 1997).

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermenters performed in a suitable medium and under conditions allowing the poly-

peptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g. in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, it can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The resulting polypeptide may be recovered by methods known in the art. For example, it may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray drying, evaporation, or precipitation.

The polypeptides may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g. ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g. preparative isoelectric focusing), differential solubility (e.g. ammonium sulfate precipitation), SDS-PAGE, or extraction (see e.g. *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989). Specific methods for purifying polypeptides exhibiting FSH activity have been described (Human Cytokines, Handbook of Basic and Clinical Research, Volume II, Blackwell Science, Eds. Aggarwal and Gutterman, 1996, pp. 19-42).

Homogeneous preparation of a conjugate of the invention

In a further aspect the invention relates to a substantially homogeneous preparation of a conjugate of the invention. In the present context a "substantially homogeneous preparation" is a preparation, typically in a suitable buffer, containing more than 50%, such as more than 75% and preferably more than 85%, or more than 90% identical conjugates, i.e. having the same degree and nature of conjugation. The substantially homogeneous preparation is conveniently obtained by ensuring that the polypeptide part of the conjugate contains the necessary number of attachment groups, located at the surface of the molecule in such a way that all attachment groups can be conjugated to the non-polypeptide moiety of choice, when the conjugation is performed in the presence of a molar excess of the non-polypeptide moiety relative to the polypeptide. Preferably, the non-polypeptide moiety to be used in this aspect of the invention is a polymer molecule.

Pharmaceutical composition of the invention and its use

In one aspect the polypeptide, the conjugate or the pharmaceutical composition according to the invention is used for the manufacture of a medicament for treatment of infertility or diseases associated with insufficient endogenous production of FSH.

In another aspect the polypeptide, the conjugate or the pharmaceutical composition according to the invention is used in a method of treating an infertile mammal, in particular a human, comprising administering to the mammal in need thereof such polypeptide, conjugate or pharmaceutical composition.

The polypeptide exhibiting FSH activity of the invention or the conjugate of the invention is administered at a dose approximately paralleling that employed in therapy with rhFSH such as Gonal-F® and Puregon®. However, due to the increased functional *in vivo* half-life of the conjugate of the invention the product should be administered less frequently and at a dose which provides a comparable effect to that obtained in current therapy. Accordingly, the exact dose to be administered depends on the circumstances, including the patient to be treated, the cause of infertility if known, the status of the ovaries, the patient's plasma FSH concentration prior to treatment, and the functional *in vivo* half-life of the product. Normally, in the

treatment of infertility the dose should be capable of stimulating follicle maturation, e.g. induce follicles to grow about 2 mm per day during a time period of 8-9 days. For instance, for a product having a functional *in vivo* half-life of 3-4 days, two doses should be given at least three days apart if a relatively stable plasma concentration is desired. Analogously, for a product having a functional *in vivo* half-life of about 6 days one dose may suffice during the entire stimulation period.

The composition of the invention may be exceedingly advantageous when employed in a step-down protocol, i.e. a protocol where decreasing dosages of FSH are given during the stimulation period, but where use of the composition may provide exactly such a slowly decreasing plasma concentration of FSH.

It will be apparent to those of skill in the art that an effective amount of a conjugate, preparation or composition of the invention depends, *inter alia*, upon the disease, the dose, the administration schedule, whether the polypeptide or conjugate or composition is administered alone or in conjunction with other therapeutic agents, the serum half-life of the compositions, and the general health of the patient. Typically, an effective dose of the conjugate, preparation or composition of the invention is sufficient to ensure development and maturation of follicles at a rate and to a degree compatible with that obtained using standard rhFSH such as Gonal-F® and Puregon®.

A further contemplated advantage is that the more stable plasma concentration obtained with a composition of the invention results in a more efficient development and maturation of follicles, which subsequently may enable a higher pregnancy rate.

The polypeptide or conjugate of the invention is preferably administered in a composition including a pharmaceutically acceptable carrier or excipient. "Pharmaceutically acceptable" means a carrier or excipient that does not cause any untoward effects in patients to whom it is administered. Such pharmaceutically acceptable carriers and excipients are well known in the art.

The polypeptide or conjugate of the invention can be formulated into pharmaceutical compositions by well-known methods. Suitable formulations are described by Remington's Pharmaceutical Sciences by E.W. Martin (Mark Publ. Co., 16th Ed., 1980).

The pharmaceutical composition of the polypeptide or conjugate of the invention may be formulated in a variety of forms, including liquids, e.g. ready-to-use solutions or suspensions, gels, lyophilized, or any other suitable form, e.g. powder or crystals suitable for preparing a solution. The preferred form will depend upon the particular indication being treated and will be apparent to one of skill in the art.

The pharmaceutical composition containing the polypeptide or conjugate of the invention may be administered intravenously, intramuscularly, intraperitoneally, intradermally, subcutaneously, sublingualy, buccally, intranasally, transdermally, by inhalation, or in any other acceptable manner, e.g. using PowderJect® or ProLease® technology or a pen injection system. The preferred mode of administration will depend upon the particular indication being treated and will be apparent to one of skill in the art. In particular, it is advantageous that the composition be administered subcutaneously, since this allows the patient to conduct the administration his-/herself.

The pharmaceutical composition of the invention may be administered in conjunction with other therapeutic agents. These agents may be incorporated as part of the same pharmaceutical composition or may be administered separately from the polypeptide or conjugate of the invention, either concurrently or in accordance with any other acceptable treatment schedule. In addition, the polypeptide, conjugate or pharmaceutical composition of the invention may be used as an adjunct to other therapies.

By obtaining a more stable FSH plasma concentration just above the threshold level for follicle growth, the composition of the invention is of particular interest for the treatment of women suffering from anovulation WHO type I, II or III, since only 1-2 mature follicles are desired in these patients.

Furthermore, the invention relates to the use of a composition of the invention in a step-down protocol where a decreasing plasma FSH concentrations are obtained using only one injection, to the use of a composition of the invention in a step-up protocol where an increase in FSH concentrations is obtained faster using a lower individual as well as total dosage, and to the use of a composition of the invention in combination with compounds for *in vitro* maturation (sterol derivatives such as FF-MAS and media containing growth and maturation factors known in the art).

Mixtures of FSH and LH activities (hMG) are routinely used in the treatment of human infertility. This particular combination therapy may be advantageous because gonadal support of gamete maturation is dependent upon the synergistic actions of both FSH and LH. Current treatment protocols requiring FSH and LH activity utilize urinary extracts from postmenopausal women. The use of these extracts is compromised by several factors, including variability.

It will in some cases be advantageous to administer the composition of the invention as part of a treatment protocol that also involves LH and/or hCG, for example recombinant LH and/or hCG. This may in particular be useful for treatment of women with low endogenous LH levels. Finally, the composition of the invention may be used, possibly in combination with LH, in the treatment of male infertility, in particular of hypogonadotrophic hypogonadism and oligo- or azoospermia. The more stable plasma concentration obtained with a composition of the invention may lead to a more efficient spermatogenesis.

The present invention will be further illustrated by the following non-limiting examples and methods.

MATERIALS AND METHODS

Sequence numbering

The amino acid sequence of hFSH-α is numbered according to the mature sequence shown in SEQ ID NO 2; an (a) suffix herein indicates the α chain. The amino acid sequence of hFSH-β is numbered according to the mature sequence shown in SEQ ID NO 4; a (b) suffix herein indicates the β chain.

Structures

HFSH-α is identical to the α chain of Human Chorionic Gonadotropin (HCG) for which two published structures are available: Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., Hen-

drickson, W. A.: *Structure* **2** pp. 545 (1994) and Lapthorn, A. J., Harris, D. C., Littlejohn, A., Lustbader, J. W., Canfield, R. E., Machin, K. J., Morgan, F. J., Isaacs, N. W.: *Nature* **369** pp. 455 (1994), both including the β chain of HCG. The β chain of hFSH is 32 percent identical to the amino acid sequence of the structural part of the β chain of HCG (see the sequence alignment of Figure 1). A series of 50 models of the 3D structure of FSH was build based on the above two available hCG structures and based on the sequence alignment in Figure 1 using the program Modeller 98 (MSI INC, 1999). The four N-terminal residues (A1(a), P2(a), D3(a) and V4(a) as well as the three C-terminal residues (H90(a), K91(a) and S92(a) were not modelled as they are not identified in the HCG structures. All of the hFSH-β chain was modelled, even the part which has no homologous residues in the HCG structures.

Accessible Surface Area (ASA)

The computer program Access (B. Lee and F.M.Richards, J. Mol.Biol. 55: 379-400 (1971)) version 2 (Copyright (c) 1983 Yale University) was used to compute the accessible surface area (ASA) of the individual atoms in the structure. This method typically uses a probe-size of 1.4Å and defines the Accessible Surface Area (ASA) as the area formed by the centre of the probe. Prior to this calculation all water molecules and all hydrogen atoms should be removed from the coordinate set, as should other atoms not directly related to the protein.

Fractional ASA of side chain

The fractional ASA of the side chain atoms is computed by division of the sum of the ASA of the atoms in the side chain with a value representing the ASA of the side chain atoms of that residue type in an extended Ala-x-Ala tripeptide, see Hubbard, Campbell & Thornton (1991) J.Mol.Biol.220,507-530. For this example the CA atom is regarded as being a part of the side chain of glycine residues but not other residues. The following values are used as standard 100% ASA for the side chain:

Ala	69.23	Å^2
Arg	200.35	Å^2
Asn	106.25	Å^2
Asp	102.06	Å^2
Cys	96.69	Å^2
Gln	140.58	Å^2
Glu	134.61	Å^2
Gly	32.28	Å^2
His	147.00	Å^2
Ile	137.91	Å^2
Leu	140.76	Å^2
Lys	162.50	Å^2
Met	156.08	Å^2
Phe	163.90	Å^2
Pro	119.65	Å^2
Ser	78.16	Å^2
Thr	101.67	Å^2
Trp	210.89	Å^2
Tyr	176.61	Å^2
Val	114.14	Å^2

Determination of surface exposed residues from structural models:
Surface accessibility and fractional ASA of side chains were calculated for each of the 50 model structures. The average value over the structural ensemble was used in the following. The N- and C-terminal residues of the FSH-α chain not included in the model are defined as having 100% side chain accessibility.

The following amino acid residues in hFSH-α and hFSH-β, respectively, have more than 25% of their side chain exposed to the surface:
A1(a), P2(a), D3(a), V4(a), Q5(a), D6(a), P8(a), E9(a), T11(a), L12(a), Q13(a), E14(a), P16(a), F17(a), Q20(a), P21(a), G22(a), A23(a), P24(a), L26(a), M29(a), F33(a), R42(a), S43(a), K44(a), K45(a), T46(a), L48(a), V49(a), Q50(a), N52(a), V61(a), K63(a), S64(a), Y65(a), N66(a), R67(a), V68(a), T69(a), M71(a), G72(a), G73(a), F74(a), K75(a), N78(a), T80(a), A81(a), H83(a), C84(a), S85(a), T86(a), Y88(a), Y89(a), H90(a), K91(a), S92(a), N1(b), S2(b), E4(b), L5(b), T6(b), N7(b), I8(b), T9(b), K14(b), E15(b), E16(b), R18(b), F19(b), I21(b), S22(b), N24(b), Y31(b), Y33(b), R35(b), D36(b), L37(b), Y39(b), K40(b), D41(b), P42(b), A43(b), R44(b), P45(b), K46(b), I47(b), K49(b), K54(b), E55(b), L56(b), V57(b), Y58(b), E59(b), T60(b), V61(b), R62(b), P64(b), G65(b), A67(b), H68(b), H69(b), D71(b), L73(b), Y74(b), T75(b), T80(b), Q81(b), H83(b), G85(b), K86(b), D88(b), S89(b), D90(b), S91(b), D93(b), T95(b), V96(b), R97(b), G98(b), L99(b), G100(b), Y103(b), S105(b), F106(b), G107(b), E108(b), M109(b), K110(b), and E111(b).

The following amino acid residues have more than 50% of their side chain exposed to the surface:
A1(a), P2(a), D3(a), V4(a), Q5(a), D6(a), P8(a), E9(a), T11(a), Q13(a), E14(a), P16(a), F17(a), Q20(a), P21(a), G22(a), A23(a), K45(a), T46(a), L48(a), V49(a), Q50(a), N52(a), K63(a), S64(a), N66(a), R67(a), T69(a), G72(a), G73(a), K75(a), T86(a), Y89(a), H90(a), K91(a), S92(a), N1(b), N7(b), T9(b), E15(b), E16(b), R18(b), F19(b), N24(b), Y33(b), D41(b), P42(b), A43(b), R44(b), P45(b), K46(b), I47(b), K54(b), E55(b), V57(b), Y58(b), E59(b), R62(b), P64(b), G65(b), A67(b), H68(b), H69(b), D71(b), L73(b), T75(b), Q81(b), H83(b), K86(b), D88(b), S89(b), D90(b), S91(b), T95(b), R97(b), G98(b), L99(b), G100(b), Y103(b), S105(b), F106(b), G107(b), E108(b), M109(b), K110(b), and E111(b).

Determining distances between atoms
The distance between atoms is most easily determined using molecular graphics software, e.g. InsightII v. 98.0, MSI Inc.

Methods used to determine the *in vitro* and *in vivo* activity of rhFSH and variants thereof

In vitro bioactivity
The *in vitro* bioactivity of conjugates or polypeptides of the invention exhibiting FSH activity may be determined by an FSH receptor activation assay. A suitable assay is the CHO-luc assay described by Chappel et al., Human Reproduction, 1998, 13(3), pp 18-35. In brief, a culture of CHO cells expressing human FSH receptor (Kelton et al., 1992, Mol. Cell. Endocribol., 89, 141-151) and firefly luciferase is incubated with the polypeptide or conjugate to be tested, and the luminescence signal is measured by use of a Packard TopCounter or a similar luminescence reader.

The bioactivity of the conjugates or polypeptides of the invention may also be measured using the CHO cell line expressing the hFSH receptor by determining the ability of the polypeptide or conjugate to elicit cAMP, using a standard cAMP assay, for instance SPA-based.

Alternatively, *in vitro* bioactivity may be determined by incubating Y1 cells expressing the FSH receptor with the polypeptide or conjugate as described by Chappel et al., *op cit*. FSH receptor activation results in an increased production of progesterone, which can be measured by radioimmuno-assay, and a dose-response relationship is established between the amount of FSH added to the Y1 cells and progesterone release.

Alternatively, the ability of a polypeptide or conjugate of the invention to compete for the binding sites with hFSH is analyzed by incubating with a labeled FSH analog, for instance biotinylated hFSH or radioiodinated hFSH.

The extracellular domains of the hFSH receptor can optionally be coupled to Fc and immobilized in 96 well plates. RhFSH or variants thereof are subsequently added and the binding of these detected using either specific anti-hFSH antibodies or biotinylated or radioiodinated hFSH.

Measurement of the in vivo half-life of conjugated and unconjugated rhFSH and variants thereof

Measurement of functional *in vivo* half-life can be carried out in a number of ways as described in the literature. For instance, the ability of the conjugates or polypeptides of the invention given once to a laboratory animal to continue to stimulate the maturation of follicles may be detected with e.g. ultrasound equipment and compared to rhFSH. An indirect measure would be to test the FSH bioactivity of plasma samples drawn at different timepoints from animals treated with the subject of the invention or rhFSH. The bioactivity could be measured using the above mentioned *in vitro* assays.

Determination of the molecular size of hFSH and variants thereof

The molecular weight of a conjugate or polypeptide of the invention is determined by SDS-PAGE, gel filtration, matrix assisted laser desorption mass spectrometry or equilibrium centrifugation

Methods for PEGylation of hFSH and variants thereof

PEGylation in microtiter plates of a tagged polypeptide with FSH activity

The polypeptide exhibiting FSH activity is expressed with a suitable tag, e.g. any of the tags exemplified in the general description above and transferring culture broth to one or more wells in a microtiter plate capable of immobilising the tagged polypeptide. When the tag is Met-Lys-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-His-Gln-Gln, a nickel-nitrilotriacetic acid (Ni-NTA) HisSorb microtiter plate commercially available from QiaGen can be used.

After allowing for immobilisation of the tagged polypeptide to the microtiter plate, the wells are washed in a buffer suitable for binding and subsequent PEGylation followed by incubating the wells with the activated PEG of choice. As an example, M-SPA-5000 from Shearwater Polymers may be used. The molar ratio of activated PEG to polypeptide should be optimised, but will typically be greater than 10:1 more typically greater than 100:1. After a suitable reaction time at ambient temperature, typically around 1 hour, the reaction is stopped by removal of the activated PEG solution. The conjugated protein is eluted from the plate by incubation with a suitable buffer. Suitable elution buffers may contain imidazole, excess NTA or another chelating compound. The conjugated protein is assayed for biological activity and immunogenicity as appropriate. The tag may optionally be cleaved off using a method known in the art, e.g. using diaminopeptidase, the Gln in pos -1 being converted to pyroglutamyl

with GCT (glutamylcyclotransferase) and finally cleaved off with PGAP (pyro-glutamyl-aminopeptidase), giving the protein. The process involves several steps of metal chelate affinity chromatography. Alternatively, the tagged polypeptide may be conjugated.

PEGylation of a polypeptide exhibiting FSH activity and having a blocked receptor-binding site

The following method can be used to optimize PEGylation of hFSH in a manner excluding PEGylation of lysines involved in receptor recognition.

A homodimer complex consisting of an FSH polypeptide and the soluble domain of the FSH receptor in a 1:1 stoichiometry is formed in a PBS buffer at pH 7. The concentration of FSH polypeptide is approximately 20 μg/ml or 1 uM and the receptor is present at equimolar concentration.

M-SPA-5000 from Shearwater Polymers, Inc. is added at 3 different concentration levels corresponding to a 5, 20 and 100 fold molar excess of FSH polypeptide. The reaction time is 30 min at RT. After the 30 min reaction period, the pH of the reaction mixture is adjusted to 2.0 and the reaction mixture is applied to a Vydac C18 column and eluted with an acetonitrile gradient essentially as described (Utsumi et al., J. Biochem., vol. 101, 1199-1208, 1987). Alternatively, and more elegantly, an isopropanol gradient can be used.

Fractions are analyzed using the primary screening assay described herein and active PEGylated FSH polypeptide obtained by this method is stored at -80°C in PBS, pH 7 containing 1 mg/ml human serum albumin (HSA).

Strategy for preparing a conjugate of the invention comprising PEG
rhFSH as well as all possible muteins of FSH comprising a single lysine to arginine substitution are prepared and characterized with respect to specific activity as compared to rhFSH to establish which, if any, lysines are critical for activity of the molecule and which may be substituted by arginine with an acceptable retention of activity.

Subsequently, rhFSH and muteins thereof, namely muteins with inserted and/or deleted lysines, are subjected to PEGylation by providing a surplus of SPA-PEG according to the procedure disclosed in WO 97/03106. Next, the specific activity of these variants is measured. Muteins permitting PEGylation with retention of acceptable activity are chosen for further work.

The above strategy may be repeated with any other attachment group, for example acidic residue substitution and suitable PEGylation chemistry. Muteins permitting PEGylation with retention of acceptable activity are chosen for further work.

The selected muteins are subjected to PEGylation with SPA-PEG according to WO 97/03106 (or another suitable PEGylation chemistry for the chosen attachment group) while varying the molecular weight of the SPA-PEG. These molecules are controlled for continued retention of acceptable activity and subjected to characterization with respect to *in vivo* half-life according to the above protocol of the Materials and Methods section. Muteins with an increased *in vivo* half-life are selected and exemplify the invention disclosed and claimed herein.

EXAMPLE 1

Extension of the N-terminus of the FSH-α subunit with additional glycosylation sites

Construction of expression plasmids

A gene encoding the human FSH-α subunit was constructed by assembly of synthetic oligonucleotides using PCR. The codon usage of the gene was optimised for high expression in mammalian cells. Furthermore, in order to achieve high gene expression, an intron (from pCI-Neo (Promega)) was included in the 5' untranslated region of the gene. The synthetic gene was subcloned behind the CMV promoter in pcDNA3.1/Hygro (Invitrogen). The sequence of the resulting plasmid, termed pBvdH977, is given in Figure 2 (FSH-α-coding sequence at position 1225 to 1572). Similarly, a synthetic gene encoding the wildtype human FSH-β subunit was constructed. Also in this construct codon usage was optimised for high expression and an intron was included in the recipient vector (pcDNA3.1/Zeo (Invitrogen)). The sequence of the resulting FSH-β-containing plasmid, termed pBvdH1022, is given in Figure 3 (FSH-β-coding sequence at position 1231 to 1617). A construct containing a modified form of FSH-α having two additional sites at its N-terminus was generated by PCR. A DNA fragment encoding the sequence Ala-Asn-Ile-Thr-Val-Asn-Ile-Thr-Val was inserted immediately upstream of the mature FSH-α sequence in pBvdH977. The sequence of the resulting plasmid, termed pBvdH1163, is given in Figure 4 (modified FSH-α–coding sequence at position 1225 to 1599).

Expression of wildtype FSH and an N-terminally α-modified form in CHO cells

For expression of wildtype FSH, 6.25 μg of pBvdH977 and 6.25 μg of pBvdH1022 were co-transfected into Chinese Hamster Ovary (CHO) K1 cells (ATCC, CCL 61) using Lipofectamine 2000 (Life Technologies) according to the manufacturer's specifications. 40-48 hrs after transfection, culture media were collected for analysis in Western blot. For expression of the modified form of FSH containing two additional glycosylation sites at the N-terminus of the α subunit, 6.25 μg of pBvdH1163 and 6.25 μg of pBvdH1022 were co-transfected into CHO K1, and culture media were collected 48 hrs after transfection, as for wildtype FSH.

Analysis of wildtype FSH and an N-terminally α-modified form by Western blotting

The FSH content of samples was analysed by Western blotting: Proteins were separated by SDS-PAGE, and a Western blot was performed using rabbit anti human FSH (AHP519, Serotec) as primary antibody, and an ImmunoPure Ultra Sensitive ABC Peroxidase Staining Kit (Pierce) for detection. FSH forms in the 1163+1022-derived sample migrated more slowly than the wildtype in the 977+1022-derived samples. This indicated that introduction of glycosylation sites at the N-terminus of the α subunit indeed leads to hyperglycosylation of FSH.

CLAIMS

1. A polypeptide conjugate exhibiting FSH activity, comprising
i) a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits differs from the corresponding wildtype subunit in that at least one amino acid residue acid residue comprising an attachment group for a non-polypeptide moiety has been introduced or removed, and
ii) a non-polypeptide moiety bound to an attachment group of said polypeptide.

2. The conjugate according to claim 1, wherein the amino acid sequence of at least one of said FSH-α and FSH-β subunits differs from that of the corresponding wildtype subunit in that an amino acid residue comprising an attachment group for the non-polypeptide moiety has been removed from the sequence.

3. The conjugate according to claim 1 or 2, wherein the amino acid sequence of at least one of said FSH-α and FSH-β subunits differs from that of the corresponding wildtype subunit in that an amino acid residue comprising an attachment group for the non-polypeptide moiety has been introduced into the sequence.

4. The conjugate according to any of claims 1-3, wherein the amino acid sequence of FSH-α differs from that of the corresponding wildtype subunit.

5. The conjugate according to any of claims 1-3, wherein the amino acid sequence of FSH-β differs from that of the corresponding wildtype subunit.

6. The conjugate according to any of claims 1-5, wherein the corresponding wildtype subunit is hFSH-α and/or hFSH-β.

7. The conjugate according to any of claims 1-6, wherein the non-polypeptide moiety is a polymer molecule.

8. The conjugate according to any of claims 1-7, wherein the polymer molecule is polyethylene glycol.

9. The conjugate according to any of claims 1-8, wherein the amino acid residue comprising an attachment group for the non-polypeptide moiety is selected from the group consisting of a lysine, asparagine, aspartic acid, glutamic acid, tyrosine and cysteine residue, preferably a lysine residue.

10. The conjugate according to claim 9, which comprises a modified FSH-α having an amino acid sequence which differs from that of hFSH-α in the removal of at least one lysine residue selected from the group consisting of K44(a), K45(a), K51(a), K63(a), K75(a), and K91(a).

11. The conjugate according to claim 9 or 10, which comprises a modified FSH-β having an amino acid sequence which differs from that of hFSH-β in the removal of at least one lysine residue selected from the group consisting of K14(b), K40(b), K46(b), K49(b), K54(b), K86(b), and K110(b).

12. The conjugate according to any of claims 9-11, wherein the modified FSH-α and modified FSH-β subunit differ from the corresponding hFSH subunit in at least one of K45(a), K63(a), K75(a), and K91(a), and at least one of K46(b), K54(b), K86(b), and K110(b).

13. The conjugate according to any of claims 6-12, wherein the polypeptide is glycosylated.

14. The conjugate according to claim 13, wherein the amino acid sequence of at least one of FSH-α and FSH-β differs from that of the corresponding wildtype sequence in that an N-glycosylation site has been introduced and/or removed.

15. A polypeptide conjugate exhibiting FSH activity comprising
i) a polypeptide comprising FSH-α and FSH-β subunits, wherein the amino acid sequence of at least one of said FSH-α and FSH-β subunits differs from that of the corresponding wildtype subunit in that at least one N-glycosylation site has been introduced, and
ii) an oligosaccharide moiety bound to an N-glycosylation site of said polypeptide.

16. The conjugate according to claim 15, wherein the amino acid sequence of at least one of said FSH-α and FSH-β subunits further differs from that of the corresponding wildtype subunit in at least one naturally-occurring N-glycosylation site has been removed.

17. The conjugate according to any of claims 13-16, wherein an N-glycosylation site has been introduced by a mutation selected from the group consisting of P2(a)N+V4(a)S, P2(a)N+V4(a)T, D3(a)N+Q5(a)S, D3(a)N+Q5(a)T, V4(a)N+D6(a)S, V4(a)N+D6(a)S, D6(a)N+P8(a)S, D6(a)N+P8(a)T, E9(a)N+T11(a)S, E9(a)N, T11(a)N+Q13(a)S, T11(a)N+Q13(a)T, L12(a)N+E14(a)S, L12(a)N+E14(a)T, E14(a)N+P16(a)S, E14(a)N+P16(a)T, P16(a)N+F18(a)S, P16(a)N+F18(a)T, F17(a)N, F17(a)N+S19(a)T, G22(a)N+P24(a)S, G22(a)N+P24(a)T, P24(a)N+L26(a)S, P24(a)N+L26(a)T, F33(a)N+R35(a)S, F33(a)N+R35(a)T, R42(a)N+K44(a)S, R42(a)N+K44(a)T, S43(a)N+K45(a)S, S43(a)N+K45(a)T, K44(a)N+T46(a)S, K44(a)N, K45(a)N+M47(a)S, K45(a)N+M47(a)T, T46(a)N+L48(a)S, T46(a)N+L48(a)T, L48(a)N+Q50(a)S, I48(a)N+Q50(a)T, V49(a)N+K51(a)S, V49(a)N+K51(a)T, Q50(a)N+N52(a)S, Q50(a)N+N52(a)T, V61(a)N+K63(a)S, V61(a)N+K63(a)T, K63(a)N+Y65(a)S, K63(a)N+Y65(a)T, S64(a)N+N66(a)S, S64(a)N+N66(a)T, Y65(a)N+R67(a)S, Y65(a)N+R67(a)T, V68(a)S, V68(a)T, R67(a)N+T69(a)S, R67(a)N, T69(a)N+M71(a)S, T69(a)N+M71(a)T, M71(a)N+G73(a)S, M71(a)N+G73(a)T, G72(a)N+F74(a)S, G72(a)N+F74(a)T, G73(a)N+K75(a)S, G73(a)N+K75(a)T, F74(a)N+V76(a)S, F74(a)N+V76(a)T, K75(a)N+E77(a)S, K75(a)N+E77(a)T, A81(a)N+H83(a)S, A81(a)N+H83(a)T, H83(a)N, T86(a)N+Y88(a)S, T86(a)N+Y88(a)T, Y88(a)N+H90(a)S, Y88(a)N+H90(a)T, Y89(a)N+K91(a)S, Y89(a)N+K91(a)T, H90(a)N and H90(a)N+S92(a)T.

18. The conjugate according to any of claims 13-17, comprising a modified FSH-β having an amino acid sequence which differs from that of hFSH-β in the introduction of at least one N-glycosylation site by a mutation selected from the group consisting of S2(b)N+E4(b)S, S2(b)N+E4(b)T, E4(b)N+T6(b)S, E4(b)N, L5(b)N+N7(b)S, L5(b)N+L7(b)T, T6(b)N+I8(b)S, T6(b)N+I8(b)T, I8(b)N+I10(b)S, I8(b)N+I10(b)T, T9(b)N+A11(b)S, T9(b)N+A11(b)T, K14(b)N+E16(b)S, K14(b)N+E16(b)T, F19(b)N+I21(b)S, F19(b)N+I21(b)T, I21(b)N+I23(b)S, I21(b)N+I23(b)T, S22(b)N+N24(b)S, S22(b)N+N24(b)T, Y31(b)N+Y33(b)S, Y31(b)N+Y33(b)T, Y33(b)N+R35(b)S, Y33(b)N+R35(b)T, R35(b)N+L37(b)S, R35(b)N+L37(b)T, D36(b)N+V38(b)S,

D36(b)N+V38(b)T, L37(b)N+Y39(b)S, L37(b)N+Y39(b)T, K40(b)N+P42(b)S, K40(b)N+P42(b)T, A43(b)N+P45(b)S, A43(b)N+P45(b)T, P45(b)N+I47(b)S, P45(b)N+I47(b)T, K46(b)N+Q48(b)S, K46(b)N+Q48(b)T, I47(b)N+K49(b)S, I47(b)N+K49(b)T, K54(b)N+L56(b)S, K54(b)N+L56(b)T, E55(b)N+V57(b)S, E55(b)N+V57(b)T, L56(b)N+Y58(b)S, L56(b)N+Y58(b)T, V57(b)N+E59(b)S, V57(b)N+E59(b)T, Y58(b)N+T60(b)S, Y58(b)N, E59(b)N+V61(b)S, E59(b)N+V61(b)T, T60(b)N+R62(b)S, T60(b)N+R62(b)T, R62(b)N+P64(b)S, R62(b)N+P64(b)T, G65(b)N+A67(b)S, G65(b)N+A67(b)T, A67(b)N+H69(b)S, A67(b)N+H69(b)T, H68(b)N+A70(b)S, H68(b)N+A70(b)T, H69(b)N+D71(b)S, H69(b)N+D71(b)T, D71(b)N+L73(b)S, D71(b)N+L73(b)T, L73(b)N+T75(b)S, L73(b)N, T75(b)N+P77(b)S, T75(b)N+P77(b)T, H83(b)N+G85(b)S, H83(b)N+G85(b)T, K86(b)N+D88(b)S, K86(b)N+D88(b)T, D88(b)N+D90(b)S, D88(b)N+D90(b)T, S89(b)N, S89(b)N+S91(b)T, D90(b)N+T92(b)S, D90(b)N, S91(b)N+D93(b)S, S91(b)N+D93(b)T, D93(b)N+T96(b)S, D93(b)N, T95(b)N+R97(b)S, T95(b)N+R97(b)T, V96(b)N+G98(b)S, V96(b)N+G98(b)T, R97(b)N+L99(b)S, R97(b)N+L99(b)T, L99(b)N+P101(b)S, L99(b)N+P101(b)T, Y103(b)N, Y103(b)N+S105(b)T, S105(b)N+G107(b)S, S105(b)N+G107(b)T, F106(b)N+E108(b)S, F106(b)N+E108(b)T, G107(b)N+M109(b)S, G107(b)N+M109(b)T, E108(b)N+K110(b)S, E108(b)N+K110(b)T, M109(b)N+E111(b)S, and M109(b)N+E111(b)T.

19. The conjugate according to any of claims 13-18, wherein a naturally occurring glycosylation site has been removed from FSH-α and/or FSH-β.

20. The conjugate according to any of claims 1-19, wherein the amino acid sequence of FSH-α and/or FSH-β differs in 1-15 amino acid residues from the corresponding wildtype sequence.

21. The conjugate according to any of claims 1-20, which comprises at least one further mutation in FSH-α and/or FSH-β, said mutation being neither an introduction nor a removal of an amino acid residue comprising an attachment group for the non-polypeptide moiety.

22. The conjugate according to any of claims 15-21, which further comprises a non-polypeptide moiety different from an N- or O-linked carbohydrate moiety.

23. The conjugate according to any of the preceding claims, which has reduced renal clearance as compared to hFSH.

24. The conjugate according to any of the preceding claims, which has an increased functional *in vivo* half-life and/or serum half-life as compared to hFSH.

25. The conjugate according to any of claims 1-24, comprising a sufficient number or type of non-polypeptide moieties to render the conjugate less susceptible to renal clearance than hFSH.

26. The conjugate according to claim 25, wherein at least one of the non-polypeptide moieties is a polymer molecule.

27. The conjugate according to any of claims 1-26, which has a molecular weight of at least about 67 kDa, in particular at least about 70 kDa.

28. The conjugate according to any of claims 23-27, said conjugate being according to claim 1 or 2 having an oligosaccharide moiety as the only type of non-polypeptide moiety and having at least one removed N-glycosylation site, but no introduced N-glycosylation site.

29. A substantially homogeneous preparation of a conjugate according to any of claims 1-28.

30. FSH-α which has an amino acid sequence that differs from that of the corresponding wildtype FSH-α subunit in that at least one amino acid residue comprising an attachment group for a polymer molecule has been introduced and/or removed.

31. FSH-β which has an amino acid sequence that differs from that of the corresponding wildtype FSH-β subunit in that at least one amino acid residue comprising an attachment group for a polymer molecule has been introduced and/or removed.

32. The FSH subunit according to claim 30 or 31, wherein a non-naturally occurring N-glycosylation site has been introduced.

33. The FSH subunit according to claim 32, wherein a naturally-occurring N-glycosylation site has been removed.

34. The FSH subunit according to any of claims 30-33, which is glycosylated.

35. A nucleotide sequence encoding a polypeptide according to any of claims 30-34.

36. An expression vector harbouring a nucleotide sequence according to claim 35.

37. A pair of expression vectors, each vector being capable of transfecting a eukaryotic cell, the vectors comprising nucleotide sequences encoding, respectively, FSH-α according to claim 35 and a wildtype FSH-β subunit, FSH-β according to claim 35 and a wildtype FSH-α subunit, or FSH-α according to claim 35 and FSH-β according to claim 35.

38. A host cell comprising a nucleotide sequence according to claim 35, an expression vector according to claim 36, or a pair of expression vectors according to claim 37.

39. The host cell according to claim 38, which is a eukaryotic cell.

40. The host cell according to claim 39, which is a mammalian cell.

41. A method for producing a modified FSH subunit according to any of claims 30-34, which method comprises subjecting the cell according to any of claims 38-40 comprising a nucleotide sequence encoding said modified subunit to cultivation under conditions conducive for expression of the subunit, and optionally recovering the subunit.

42. The method according to claim 41, which further comprises subjecting the subunit to conjugation to a non-polypeptide moiety so as to produce a conjugate according to any of claims 1-28 or a preparation according to claim 29.

43. The method according to claim 42, wherein the non-polypeptide moiety is a polymer molecule and the conjugation is performed in the presence of a molar excess of the polymer

moiety relative to the polypeptide, whereby a substantially homogeneous preparation of conjugates is obtained.

44. A method for increasing the functional *in vivo* half-life and/or serum half-life of a polypeptide exhibiting FSH activity, which method comprises introducing an amino acid residue change as defined in any of claims 1-28 and subjecting the resulting modified polypeptide to conjugation with an appropriate non-polypeptide moiety.

45. A method for preparing a conjugate according to any of claims 1-28, comprising providing a polypeptide i) and a non-polypeptide moiety ii), allowing the polypeptide to react with the non-polypeptide moiety under conditions conducive for conjugation to take place, and recovering the resulting conjugate.

46. The method according to any of claims 41-45, wherein conjugation to the non-polypeptide moiety is conducted in the presence of a molar excess of the non-polypeptide moiety relative to the polypeptide, whereby a substantially homogenous conjugate preparation is obtained.

47. A method for preparing a polypeptide exhibiting FSH activity comprising a modified FSH-α subunit according to any of claims 30 or 32-34 and a wildtype FSH β-subunit, a modified FSH-β subunit according to any of claims 31-34 and a wildtype FSH-α subunit, or a modified FSH-α subunit according to any of claims 30 or 32-34 and a modified FSH-β subunit according to any of claims 31-34, which method comprises producing the respective subunits separately and allowing the subunits to dimerize.

48. The method according to claim 47, which further comprises subjecting the resulting dimeric polypeptide to conjugation with a non-polypeptide moiety.

49. A pharmaceutical composition comprising a) a conjugate according to any of claims 1-28 or a preparation according to claim 29, and b) a pharmaceutically acceptable diluent, carrier or adjuvant.

50. A conjugate according to any of claims 1-28, a preparation according to claim 29, or a composition according to claim 49 for use in the treatment of infertility.

51. Use of a conjugate according to any of claims 1-28, a preparation according to claim 29, or a composition according to claim 49 for the treatment of infertility.

52. Use of a conjugate according to any of claims 1-28, a preparation according to claim 22, or a composition according to claim 49 for the manufacture of a medicament for treatment of infertility.

53. A method of treating an infertile mammal comprising administering to a mammal in need thereof an effective amount of a conjugate according to any of claims 1-28, a preparation according to claim 22, or a composition according to claim 35.

54. A polypeptide conjugate exhibiting FSH activity, comprising a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits comprises a polymer molecule bound to the N-terminal thereof.

55. The polypeptide of claim 54, wherein the polymer molecule is polyethylene glycol.

56. A polypeptide conjugate exhibiting FSH activity, comprising a polypeptide comprising FSH-α and FSH-β subunits, wherein at least one of said FSH-α and FSH-β subunits comprises at least one introduced N- or O-glycosylation site at the N-terminal thereof, said at least one introduced glycosylation site being glycosylated.

57. The polypeptide conjugate of any of claims 54-56, wherein the FSH-α subunit comprises hFSH-α having the sequence shown in SEQ ID NO 2 and/or the FSH-β subunit comprises hFSH-β having the sequence shown in SEQ ID NO 4.

58. The polypeptide conjugate of claim 54 or 55, said conjugate further being as defined in any of claims 1-27.

59. The polypeptide conjugate of claim 56, said conjugate further being as defined in any of claims 1-12 or 16-27.

SEQUENCE LISTING

SEQ ID NO 1

The complete amino acid sequence of the common α chain, named "Glycoprotein hormones α chain" Fiddes J.C., Goodman H.M. "Isolation, cloning and sequence analysis of the cDNA for the α-subunit of human chorionic gonadotropin." *Nature* **281**:351-356(1979).

MDYYRKYAAI FLVTLSVFLH VLHSAPDVQD CPECTLQENP FFSQPGAPIL QCMGCCFSRA YPTPLRSKKT MLVQKNVTSE STCCVAKSYN RVTVMGGFKV ENHTACHCST CYYHKS

Rathnam P., Saxena B.B.; "Primary amino acid sequence of follicle-stimulating hormone from human pituitary glands. I. α subunit." *J. Biol. Chem.* **250**:6735-6746(1975). Reports residue Q29 to be a Glu.

Sairam M.R., Li C.H. "Human pituitary thyrotropin. The primary structure of the α and beta subunits." *Can. J. Biochem.* **55**:755-760(1977), and Sairam M.R., Papkoff H., Li C.H. "Human pituitary interstitial cell stimulating hormone: primary structure of the α-subunit." *Biochem. Biophys. Res. Commun.* **48**:530-537(1972) report the sequence CS at positions 108-109 to be the sequence SC.

SEQ ID NO 2

The mature amino acid sequence of the common α chain shown in SEQ ID NO 1.

APDVQDCPEC TLQENPFFSQ PGAPILQCMG CCFSRAYPTP LRSKKTMLVQ KNVTSESTCC VAKSYNRVTV MGGFKVENHT ACHCSTCYYH KS

SEQ ID NO 3

The complete amino acid sequence of Human FSH β chain, Tanzi R.E., Gusella J.F., Shows T.B. "DNA sequence and regional assignment of the human follicle-stimulating hormone beta-subunit gene to the short arm of human chromosome 11." *DNA* **6**:205-212(1987).

MKTLQFFFLF CCWKAICCNS CELTNITIAI EKEECRFCIS INTTWCAGYC YTRDLVYKDP ARPKIQKTCT FKELVYETVR VPGCAHHADS LYTYPVATQC HCGKCDSDST DCTVRGLGPS YCSFGEMKE

SEQ ID NO 4

The mature sequence of Human FSH shown in SEQ ID NO 3.

NSCELTNITI AIEKEECRFC ISINTTWCAG YCYTRDLVYK DPARPKIQKT CTFKELVYET VRVPGCAHHA DSLYTYPVAT QCHCGKCDSD STDCTVRGLG PSYCSFGEMK E

FIGURE 1

Sequence alignments:

Sequence alignment of Human FSH to the structural part of the two structures of Human Chorionic Gonadotropin. The "/" indicates the chain break between the alpha and the beta chain.

```
FSH    -QDCPECTLQ ENPFFSQPGA PILQCMGCCF SRAYPTPLRS KKTMLVQKNV
1HRP   TQDCPECTLQ ENPFFSQPGA PILQCMGCCF SRAYPTPLRS KKTMLVQKNV
1HCN   -QDCPECTLQ ENPFFSQPGA PILQCMGCCF SRAYPTPLRS KKTMLVQKNV

FSH    TSESTCCVAK SYNRVTVMGG FKVENHTACH CSTCYY/--- --NSCELTNI
1HRP   TSESTCCVAK SYNRVTVMGG FKVENHTACH CSTCYY/KEP LRPRCRPINA
1HCN   TSESTCCVAK SYNRVTVMGG FKVENHTACH CSTCYY/KEP LRPRCRPINA

FSH    TIAIEKEECR FCISINTTWC AGYCYTRDLV YKDPARPKIQ KTCTFKELVY
1HRP   TLAVEKEGCP VCITVNTTIC AGYCPTMTRV LQGVLPALPQ VVCNYRDVRF
1HCN   TLAVEKEGCP VCITVNTTIC AGYCPTMTRV LQGVLPALPQ VVCNYRDVRF

FSH    ETVRVPGCAH HADSLYTYPV ATQCHCGKCD SDSTDCTVRG LGPSYCSFGE
1HRP   ESIRLPGCPR GVNPVVSYAV ALSCQCALCR RSTTDCGGPK DHPLTCD...
1HCN   ESIRLPGCPR GVNPVVSYAV ALSCQCALCR RSTTDCGGPK DHPLTCD...

FSH    MKE
1HRP   ...
1HCN   ...
```

FIGURE 2 (p. 1/5)

```
   1  GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC
      CTGCCTAGCC CTCTAGAGGG CTAGGGGATA CCAGCTGAGA GTCATGTTAG

  51  TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
      ACGAGACTAC GGCGTATCAA TTCGGTCATA GACGAGGGAC GAACACACAA

 101  GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
      CCTCCAGCGA CTCATCACGC GCTCGTTTTA AATTCGATGT TGTTCCGTTC

 151  GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG
      CGAACTGGCT GTTAACGTAC TTCTTAGACG AATCCCAATC CGCAAAACGC

 201  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT GATTATTGAC
      GACGAAGCGC TACATGCCCG GTCTATATGC GCAACTGTAA CTAATAACTG

 251  TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
      ATCAATAATT ATCATTAGTT AATGCCCCAG TAATCAAGTA TCGGGTATAT

 301  TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG
      ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC

 351  CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
      GGGTTGCTGG GGGCGGGTAA CTGCAGTTAT TACTGCATAC AAGGGTATCA

 401  AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAC TATTTACGGT
      TTGCGGTTAT CCCTGAAAGG TAACTGCAGT TACCCACCTG ATAAATGCCA

 451  AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC
      TTTGACGGGT GAACCGTCAT GTAGTTCACA TAGTATACGG TTCATGCGGG

 501  CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA
      GGATAACTGC AGTTACTGCC ATTTACCGGG CGGACCGTAA TACGGGTCAT

 551  CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA
      GTACTGGAAT ACCCTGAAAG GATGAACCGT CATGTAGATG CATAATCAGT

 601  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA
      AGCGATAATG GTACCACTAC GCCAAAACCG TCATGTAGTT ACCCGCACCT

 651  TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
      ATCGCCAAAC TGAGTGCCCC TAAAGGTTCA GAGGTGGGGT AACTGCAGTT

 701  TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA
      ACCCTCAAAC AAAACCGTGG TTTTAGTTGC CCTGAAAGGT TTTACAGCAT

 751  ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG
      TGTTGAGGCG GGGTAACTGC GTTTACCCGC CATCCGCACA TGCCACCCTC

 801  GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG
      CAGATATATT CGTCTCGAGA GACCGATTGA TCTCTTGGGT GACGAATGAC

 851  GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
      CGAATAGCTT TAATTATGCT GAGTGATATC CCTCTGGGTT CGACCGATCG

 901  TTATTGCGGT AGTTTATCAC AGTTAAATTG CTAACGCAGT CAGTGCTTCT
      AATAACGCCA TCAAATAGTG TCAATTTAAC GATTGCGTCA GTCACGAAGA

 951  GACACAACAG TCTCGAACTT AAGCTGCAGT GACTCTCTTA AGGTAGCCTT
      CTGTGTTGTC AGAGCTTGAA TTCGACGTCA CTGAGAGAAT TCCATCGGAA

1001  GCAGAAGTTG GTCGTGAGGC ACTGGGCAGG TAAGTATCAA GGTTACAAGA
      CGTCTTCAAC CAGCACTCCG TGACCCGTCC ATTCATAGTT CCAATGTTCT

1051  CAGGTTTAAG GAGACCAATA GAAACTGGGC TTGTCGAGAC AGAGAAGACT
      GTCCAAATTC CTCTGGTTAT CTTTGACCCG AACAGCTCTG TCTCTTCTGA

1101  CTTGCGTTTC TGATAGGCAC CTATTGGTCT TACTGACATC CACTTTGCCT
      GAACGCAAAG ACTATCCGTG GATAACCAGA ATGACTGTAG GTGAAACGGA

1151  TTCTCTCCAC AGGTGTCCAC TCCCAGTTCA ATTACAGCTC TTAAAAGCTT
      AAGAGAGGTG TCCACAGGTG AGGGTCAAGT TAATGTCGAG AATTTTCGAA

  +1                                  Met  Asp  Tyr  Tyr  Arg  Lys  Tyr  Ala  Ala

1201  GGTACCGAGC TCGGATCCGC CACCATGGAC TACTACCGCA AGTACGCCGC
      CCATGGCTCG AGCCTAGGCG GTGGTACCTG ATGATGGCGT TCATGCGGCG
```

FIGURE 2 (p. 2/5)

```
 -1  Ala Ile Phe Leu Val Thr Leu Ser Val Phe Leu His Val Leu His Ser Ala Pro
1251 CATCTTCCTG GTGACCCTGA GCGTGTTCCT GCACGTGCTG CACAGCGCCC
     GTAGAAGGAC CACTGGGACT CGCACAAGGA CGTGCACGAC GTGTCGCGGG

 -1  Pro Asp Val Gln Asp Cys Pro Glu Cys Thr Leu Gln Glu Asn Pro Phe Phe
1301 CCGACGTGCA GGACTGCCCC GAGTGCACCC TGCAGGAGAA CCCCTTCTTC
     GGCTGCACGT CCTGACGGGG CTCACGTGGG ACGTCCTCTT GGGGAAGAAG

 -1  Ser Gln Pro Gly Ala Pro Ile Leu Gln Cys Met Gly Cys Cys Phe Ser Arg
1351 AGCCAGCCCG GCGCCCCCAT CCTGCAGTGC ATGGGCTGCT GCTTCAGCCG
     TCGGTCGGGC CGCGGGGGTA GGACGTCACG TACCCGACGA CGAAGTCGGC

 -1  Arg Ala Tyr Pro Thr Pro Leu Arg Ser Lys Lys Thr Met Leu Val Gln Lys Asn
1401 CGCCTACCCC ACCCCCCTGC GCAGCAAGAA GACCATGCTG GTGCAGAAGA
     GCGGATGGGG TGGGGGGACG CGTCGTTCTT CTGGTACGAC CACGTCTTCT

 -1  Asn Val Thr Ser Glu Ser Thr Cys Cys Val Ala Lys Ser Tyr Asn Arg Val
1451 ACGTGACCAG CGAGAGCACC TGCTGCGTGG CCAAGAGCTA CAACCGCGTG
     TGCACTGGTC GCTCTCGTGG ACGACGCACC GGTTCTCGAT GTTGGCGCAC

 -1  Thr Val Met Gly Gly Phe Lys Val Glu Asn His Thr Ala Cys His Cys Ser
1501 ACCGTGATGG GCGGCTTCAA GGTGGAGAAC CACACCGCCT GCCACTGCAG
     TGGCACTACC CGCCGAAGTT CCACCTCTTG GTGTGGCGGA CGGTGACGTC

 -1  Ser Thr Cys Tyr Tyr His Lys
1551 CACCTGCTAC TACCACAAGA GCTAATCTAG AGGGCCCGTT TAAACCCGCT
     GTGGACGATG ATGGTGTTCT CGATTAGATC TCCCGGGCAA ATTTGGGCGA

1601 GATCAGCCTC GACTGTGCCT TCTAGTTGCC AGCCATCTGT TGTTTGCCCC
     CTAGTCGGAG CTGACACGGA AGATCAACGG TCGGTAGACA ACAAACGGGG

1651 TCCCCCGTGC CTTCCTTGAC CCTGGAAGGT GCCACTCCCA CTGTCCTTTC
     AGGGGGCACG GAAGGAACTG GGACCTTCCA CGGTGAGGGT GACAGGAAAG

1701 CTAATAAAAT GAGGAAATTG CATCGCATTG TCTGAGTAGG TGTCATTCTA
     GATTATTTTA CTCCTTTAAC GTAGCGTAAC AGACTCATCC ACAGTAAGAT

1751 TTCTGGGGGG TGGGGTGGGG CAGGACAGCA AGGGGGAGGA TTGGGAAGAC
     AAGACCCCCC ACCCCACCCC GTCCTGTCGT TCCCCCTCCT AACCCTTCTG

1801 AATAGCAGGC ATGCTGGGGA TGCGGTGGGC TCTATGGCTT CTGAGGCGGA
     TTATCGTCCG TACGACCCCT ACGCCACCCG AGATACCGAA GACTCCGCCT

1851 AAGAACCAGC TGGGGCTCTA GGGGGTATCC CCACGCGCCC TGTAGCGGCG
     TTCTTGGTCG ACCCCGAGAT CCCCCATAGG GGTGCGCGGG ACATCGCCGC

1901 CATTAAGCGC GGCGGGTGTG GTGGTTACGC GCAGCGTGAC CGCTACACTT
     GTAATTCGCG CCGCCCACAC CACCAATGCG CGTCGCACTG GCGATGTGAA

1951 GCCAGCGCCC TAGCGCCCGC TCCTTTCGCT TTCTTCCCTT CCTTTCTCGC
     CGGTCGCGGG ATCGCGGGCG AGGAAAGCGA AAGAAGGGAA GGAAAGAGCG

2001 CACGTTCGCC GGCTTTCCCC GTCAAGCTCT AAATCGGGGC ATCCCTTTAG
     GTGCAAGCGG CCGAAAGGGG CAGTTCGAGA TTTAGCCCCG TAGGGAAATC

2051 GGTTCCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG
     CCAAGGCTAA ATCACGAAAT GCCGTGGAGC TGGGGTTTTT TGAACTAATC

2101 GGTGATGGTT CACGTAGTGG GCCATCGCCC TGATAGACGG TTTTTCGCCC
     CCACTACCAA GTGCATCACC CGGTAGCGGG ACTATCTGCC AAAAAGCGGG

2151 TTTGACGTTG GAGTCCACGT TCTTTAATAG TGGACTCTTG TTCCAAACTG
     AAACTGCAAC CTCAGGTGCA AGAAATTATC ACCTGAGAAC AAGGTTTGAC

2201 GAACAACACT CAACCCTATC TCGGTCTATT CTTTTGATTT ATAAGGGATT
     CTTGTTGTGA GTTGGGATAG AGCCAGATAA GAAAACTAAA TATTCCCTAA

2251 TTGGGGATTT CGGCCTATTG GTTAAAAAAT GAGCTGATTT AACAAAAATT
     AACCCCTAAA GCCGGATAAC CAATTTTTTA CTCGACTAAA TTGTTTTTAA

2301 TAACGCGAAT TAATTCTGTG GAATGTGTGT CAGTTAGGGT GTGGAAAGTC
     ATTGCGCTTA ATTAAGACAC CTTACACACA GTCAATCCCA CACCTTTCAG

2351 CCCAGGCTCC CCAGGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG
     GGGTCCGAGG GGTCCGTCCG TCTTCATACG TTTCGTACGT AGAGTTAATC
```

```
2401  TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT
      AGTCGTTGGT CCACACCTTT CAGGGGTCCG AGGGGTCGTC CGTCTTCATA

2451  GCAAAGCATG CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC
      CGTTTCGTAC GTAGAGTTAA TCAGTCGTTG GTATCAGGGC GGGGATTGAG

2501  CGCCCATCCC GCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCCAT
      GCGGGTAGGG CGGGGATTGA GGCGGGTCAA GGCGGGTAAG AGGCGGGGTA

2551  GGCTGACTAA TTTTTTTTAT TTATGCAGAG GCCGAGGCCG CCTCTGCCTC
      CCGACTGATT AAAAAAAATA AATACGTCTC CGGCTCCGGC GGAGACGGAG

2601  TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC CTAGGCTTTT
      ACTCGATAAG GTCTTCATCA CTCCTCCGAA AAAACCTCCG GATCCGAAAA

2651  GCAAAAAGCT CCCGGGAGCT TGTATATCCA TTTTCGGATC TGATCAGCAC
      CGTTTTTCGA GGGCCCTCGA ACATATAGGT AAAAGCCTAG ACTAGTCGTG

2701  GTGATGAAAA AGCCTGAACT CACCGCGACG TCTGTCGAGA AGTTTCTGAT
      CACTACTTTT TCGGACTTGA GTGGCGCTGC AGACAGCTCT TCAAAGACTA

2751  CGAAAAGTTC GACAGCGTCT CCGACCTGAT GCAGCTCTCG GAGGGCGAAG
      GCTTTTCAAG CTGTCGCAGA GGCTGGACTA CGTCGAGAGC CTCCCGCTTC

2801  AATCTCGTGC TTTCAGCTTC GATGTAGGAG GGCGTGGATA TGTCCTGCGG
      TTAGAGCACG AAAGTCGAAG CTACATCCTC CCGCACCTAT ACAGGACGCC

2851  GTAAATAGCT GCGCCGATGG TTTCTACAAA GATCGTTATG TTTATCGGCA
      CATTTATCGA CGCGGCTACC AAAGATGTTT CTAGCAATAC AAATAGCCGT

2901  CTTTGCATCG GCCGCGCTCC CGATTCCGGA AGTGCTTGAC ATTGGGGAAT
      GAAACGTAGC CGGCGCGAGG GCTAAGGCCT TCACGAACTG TAACCCCTTA

2951  TCAGCGAGAG CCTGACCTAT TGCATCTCCC GCCGTGCACA GGGTGTCACG
      AGTCGCTCTC GGACTGGATA ACGTAGAGGG CGGCACGTGT CCCACAGTGC

3001  TTGCAAGACC TGCCTGAAAC CGAACTGCCC GCTGTTCTGC AGCCGGTCGC
      AACGTTCTGG ACGGACTTTG GCTTGACGGG CGACAAGACG TCGGCCAGCG

3051  GGAGGCCATG GATGCGATCG CTGCGGCCGA TCTTAGCCAG ACGAGCGGGT
      CCTCCGGTAC CTACGCTAGC GACGCCGGCT AGAATCGGTC TGCTCGCCCA

3101  TCGGCCCATT CGGACCGCAA GGAATCGGTC AATACACTAC ATGGCGTGAT
      AGCCGGGTAA GCCTGGCGTT CCTTAGCCAG TTATGTGATG TACCGCACTA

3151  TTCATATGCG CGATTGCTGA TCCCCATGTG TATCACTGGC AAACTGTGAT
      AAGTATACGC GCTAACGACT AGGGGTACAC ATAGTGACCG TTTGACACTA

3201  GGACGACACC GTCAGTGCGT CCGTCGCGCA GGCTCTCGAT GAGCTGATGC
      CCTGCTGTGG CAGTCACGCA GGCAGCGCGT CCGAGAGCTA CTCGACTACG

3251  TTTGGGCCGA GGACTGCCCC GAAGTCCGGC ACCTCGTGCA CGCGGATTTC
      AAACCCGGCT CCTGACGGGG CTTCAGGCCG TGGAGCACGT GCGCCTAAAG

3301  GGCTCCAACA ATGTCCTGAC GGACAATGGC CGCATAACAG CGGTCATTGA
      CCGAGGTTGT TACAGGACTG CCTGTTACCG GCGTATTGTC GCCAGTAACT

3351  CTGGAGCGAG GCGATGTTCG GGGATTCCCA ATACGAGGTC GCCAACATCT
      GACCTCGCTC CGCTACAAGC CCCTAAGGGT TATGCTCCAG CGGTTGTAGA

3401  TCTTCTGGAG GCCGTGGTTG GCTTGTATGG AGCAGCAGAC GCGCTACTTC
      AGAAGACCTC CGGCACCAAC CGAACATACC TCGTCGTCTG CGCGATGAAG

3451  GAGCGGAGGC ATCCGGAGCT TGCAGGATCG CCGCGGCTCC GGGCGTATAT
      CTCGCCTCCG TAGGCCTCGA ACGTCCTAGC GGCGCCGAGG CCCGCATATA

3501  GCTCCGCATT GGTCTTGACC AACTCTATCA GAGCTTGGTT GACGGCAATT
      CGAGGCGTAA CCAGAACTGG TTGAGATAGT CTCGAACCAA CTGCCGTTAA

3551  TCGATGATGC AGCTTGGGCG CAGGGTCGAT GCGACGCAAT CGTCCGATCC
      AGCTACTACG TCGAACCCGC GTCCCAGCTA CGCTGCGTTA GCAGGCTAGG

3601  GGAGCCGGGA CTGTCGGGCG TACACAAATC GCCCGCAGAA GCGCGGCCGT
      CCTCGGCCCT GACAGCCCGC ATGTGTTTAG CGGGCGTCTT CGCGCCGGCA

3651  CTGGACCGAT GGCTGTGTAG AAGTACTCGC CGATAGTGGA AACCGACGCC
      GACCTGGCTA CCGACACATC TTCATGAGCG GCTATCACCT TTGGCTGCGG
```

```
3701  CCAGCACTCG TCCGAGGGCA AAGGAATAGC ACGTGCTACG AGATTTCGAT
      GGTCGTGAGC AGGCTCCCGT TTCCTTATCG TGCACGATGC TCTAAAGCTA

3751  TCCACCGCCG CCTTCTATGA AAGGTTGGGC TTCGGAATCG TTTTCCGGGA
      AGGTGGCGGC GGAAGATACT TTCCAACCCG AAGCCTTAGC AAAAGGCCCT

3801  CGCCGGCTGG ATGATCCTCC AGCGCGGGGA TCTCATGCTG GAGTTCTTCG
      GCGGCCGACC TACTAGGAGG TCGCGCCCCT AGAGTACGAC CTCAAGAAGC

3851  CCCACCCCAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT
      GGGTGGGGTT GAACAAATAA CGTCGAATAT TACCAATGTT TATTTCGTTA

3901  AGCATCACAA ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG
      TCGTAGTGTT TAAAGTGTTT ATTTCGTAAA AAAAGTGACG TAAGATCAAC

3951  TGGTTTGTCC AAACTCATCA ATGTATCTTA TCATGTCTGT ATACCGTCGA
      ACCAAACAGG TTTGAGTAGT TACATAGAAT AGTACAGACA TATGGCAGCT

4001  CCTCTAGCTA GAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA
      GGAGATCGAT CTCGAACCGC ATTAGTACCA GTATCGACAA AGGACACACT

4051  AATTGTTATC CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA
      TTAACAATAG GCGAGTGTTA AGGTGTGTTG TATGCTCGGC CTTCGTATTT

4101  GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTCCGT
      CACATTTCGG ACCCCACGGA TTACTCACTC GATTGAGTGT AATTAACGCA

4151  TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
      ACGCGAGTGA CGGGCGAAAG GTCAGCCCTT TGGACAGCAC GGTCGACGTA

4201  TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC
      ATTACTTAGC CGGTTGCGCG CCCCTCTCCG CCAAACGCAT AACCCGCGAG

4251  TTCCGCTTCC TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC
      AAGGCGAAGG AGCGAGTGAC TGAGCGACGC GAGCCAGCAA GCCGACGCCG

4301  GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA
      CTCGCCATAG TCGAGTGAGT TTCCGCCATT ATGCCAATAG GTGTCTTAGT

4351  GGGGATAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG
      CCCCTATTGC GTCCTTTCTT GTACACTCGT TTTCCGGTCG TTTTCCGGTC

4401  GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC
      CTTGGCATTT TTCCGGCGCA ACGACCGCAA AAAGGTATCC GAGGCGGGGG

4451  CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG
      GACTGCTCGT AGTGTTTTTA GCTGCGAGTT CAGTCTCCAC CGCTTTGGGC

4501  ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG
      TGTCCTGATA TTTCTATGGT CCGCAAAGGG GGACCTTCGA GGGAGCACGC

4551  CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC
      GAGAGGACAA GGCTGGGACG GCGAATGGCC TATGGACAGG CGGAAAGAGG

4601  CTTCGGGAAG CGTGGCGCTT TCTCAATGCT CACGCTGTAG GTATCTCAGT
      GAAGCCCTTC GCACCGCGAA AGAGTTACGA GTGCGACATC CATAGAGTCA

4651  TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCGT
      AGCCACATCC AGCAAGCGAG GTTCGACCCG ACACACGTGC TTGGGGGGCA

4701  TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC
      AGTCGGGCTG GCGACGCGGA ATAGGCCATT GATAGCAGAA CTCAGGTTGG

4751  CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT
      GCCATTCTGT GCTGAATAGC GGTGACCGTC GTCGGTGACC ATTGTCCTAA

4801  AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC
      TCGTCTCGCT CCATACATCC GCCACGATGT CTCAAGAACT TCACCACCGG

4851  TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA
      ATTGATGCCG ATGTGATCTT CCTGTCATAA ACCATAGACG CGAGACGACT

4901  AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA
      TCGGTCAATG GAAGCCTTTT TCTCAACCAT CGAGAACTAG GCCGTTTGTT

4951  ACCACCGCTG GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC AGATTACGCG
      TGGTGGCGAC CATCGCCACC AAAAAAACAA ACGTTCGTCG TCTAATGCGC
```

```
5001  CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG
      GTCTTTTTTT CCTAGAGTTC TTCTAGGAAA CTAGAAAAGA TGCCCCAGAC

5051  ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA
      TGCGAGTCAC CTTGCTTTTG AGTGCAATTC CCTAAAACCA GTACTCTAAT

5101  TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA
      AGTTTTTCCT AGAAGTGGAT CTAGGAAAAT TTAATTTTTA CTTCAAAATT

5151  ATCAATCTAA AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT
      TAGTTAGATT TCATATATAC TCATTTGAAC CAGACTGTCA ATGGTTACGA

5201  TAATCAGTGA GGCACCTATC TCAGCGATCT GTCTATTTCG TTCATCCATA
      ATTAGTCACT CCGTGGATAG AGTCGCTAGA CAGATAAAGC AAGTAGGTAT

5251  GTTGCCTGAC TCCCCGTCGT GTAGATAACT ACGATACGGG AGGGCTTACC
      CAACGGACTG AGGGGCAGCA CATCTATTGA TGCTATGCCC TCCCGAATGG

5301  ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC
      TAGACCGGGG TCACGACGTT ACTATGGCGC TCTGGGTGCG AGTGGCCGAG

5351  CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT
      GTCTAAATAG TCGTTATTTG GTCGGTCGGC CTTCCCGGCT CGCGTCTTCA

5401  GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA
      CCAGGACGTT GAAATAGGCG GAGGTAGGTC AGATAATTAA CAACGGCCCT

5451  AGCTAGAGTA AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA
      TCGATCTCAT TCATCAAGCG GTCAATTATC AAACGCGTTG CAACAACGGT

5501  TTGCTACAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT GGCTTCATTC
      AACGATGTCC GTAGCACCAC AGTGCGAGCA GCAAACCATA CCGAAGTAAG

5551  AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC CCATGTTGTG
      TCGAGGCCAA GGGTTGCTAG TTCCGCTCAA TGTACTAGGG GGTACAACAC

5601  CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC AGAAGTAAGT
      GTTTTTTCGC CAATCGAGGA AGCCAGGAGG CTAGCAACAG TCTTCATTCA

5651  TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT
      ACCGGCGTCA CAATAGTGAG TACCAATACC GTCGTGACGT ATTAAGAGAA

5701  ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC
      TGACAGTACG GTAGGCATTC TACGAAAAGA CACTGACCAC TCATGAGTTG

5751  CAAGTCATTC TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG
      GTTCAGTAAG ACTCTTATCA CATACGCCGC TGGCTCAACG AGAACGGGCC

5801  CGTCAATACG GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGCTC
      GCAGTTATGC CCTATTATGG CGCGGTGTAT CGTCTTGAAA TTTTCACGAG

5851  ATCATTGGAA AACGTTCTTC GGGGCGAAAA CTCTCAAGGA TCTTACCGCT
      TAGTAACCTT TTGCAAGAAG CCCCGCTTTT GAGAGTTCCT AGAATGGCGA

5901  GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG
      CAACTCTAGG TCAAGCTACA TTGGGTGAGC ACGTGGGTTG ACTAGAAGTC

5951  CATCTTTTAC TTTCACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA
      GTAGAAAATG AAAGTGGTCG CAAAGACCCA CTCGTTTTTG TCCTTCCGTT

6001  AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT
      TTACGGCGTT TTTTCCCTTA TTCCCGCTGT GCCTTTACAA CTTATGAGTA

6051  ACTCTTCCTT TTTCAATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA
      TGAGAAGGAA AAAGTTATAA TAACTTCGTA AATAGTCCCA ATAACAGAGT

6101  TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA AATAGGGGTT
      ACTCGCCTAT GTATAAACTT ACATAAATCT TTTTATTTGT TTATCCCCAA

6151  CCGCGCACAT TTCCCCGAAA AGTGCCACCT GACGTC
      GGCGCGTGTA AAGGGGCTTT TCACGGTGGA CTGCAG
```

FIGURE 3 (p. 1/5)

```
   1  GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC
      CTGCCTAGCC CTCTAGAGGG CTAGGGGATA CCAGCTGAGA GTCATGTTAG

  51  TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
      ACGAGACTAC GGCGTATCAA TTCGGTCATA GACGAGGGAC GAACACACAA

 101  GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
      CCTCCAGCGA CTCATCACGC GCTCGTTTTA AATTCGATGT TGTTCCGTTC

 151  GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG
      CGAACTGGCT GTTAACGTAC TTCTTAGACG AATCCCAATC CGCAAAACGC

 201  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT GATTATTGAC
      GACGAAGCGC TACATGCCCG GTCTATATGC GCAACTGTAA CTAATAACTG

 251  TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
      ATCAATAATT ATCATTAGTT AATGCCCCAG TAATCAAGTA TCGGGTATAT

 301  TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG
      ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC

 351  CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
      GGGTTGCTGG GGGCGGGTAA CTGCAGTTAT TACTGCATAC AAGGGTATCA

 401  AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAC TATTTACGGT
      TTGCGGTTAT CCCTGAAAGG TAACTGCAGT TACCCACCTG ATAAATGCCA

 451  AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC
      TTTGACGGGT GAACCGTCAT GTAGTTCACA TAGTATACGG TTCATGCGGG

 501  CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA
      GGATAACTGC AGTTACTGCC ATTTACCGGG CGGACCGTAA TACGGGTCAT

 551  CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA
      GTACTGGAAT ACCCTGAAAG GATGAACCGT CATGTAGATG CATAATCAGT

 601  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA
      AGCGATAATG GTACCACTAC GCCAAAACCG TCATGTAGTT ACCCGCACCT

 651  TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
      ATCGCCAAAC TGAGTGCCCC TAAAGGTTCA GAGGTGGGGT AACTGCAGTT

 701  TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA
      ACCCTCAAAC AAAACCGTGG TTTTAGTTGC CCTGAAAGGT TTTACAGCAT

 751  ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG
      TGTTGAGGCG GGGTAACTGC GTTTACCCGC CATCCGCACA TGCCACCCTC

 801  GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG
      CAGATATATT CGTCTCGAGA GACCGATTGA TCTCTTGGGT GACGAATGAC

 851  GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
      CGAATAGCTT TAATTATGCT GAGTGATATC CCTCTGGGTT CGACCGATCG

 901  TTATTGCGGT AGTTTATCAC AGTTAAATTG CTAACGCAGT CAGTGCTTCT
      AATAACGCCA TCAAATAGTG TCAATTTAAC GATTGCGTCA GTCACGAAGA

 951  GACACAACAG TCTCGAACTT AAGCTGCAGT GACTCTCTTA AGGTAGCCTT
      CTGTGTTGTC AGAGCTTGAA TTCGACGTCA CTGAGAGAAT TCCATCGGAA

1001  GCAGAAGTTG GTCGTGAGGC ACTGGGCAGG TAAGTATCAA GGTTACAAGA
      CGTCTTCAAC CAGCACTCCG TGACCCGTCC ATTCATAGTT CCAATGTTCT

1051  CAGGTTTAAG GAGACCAATA GAAACTGGGC TTGTCGAGAC AGAGAAGACT
      GTCCAAATTC CTCTGGTTAT CTTTGACCCG AACAGCTCTG TCTCTTCTGA

1101  CTTGCGTTTC TGATAGGCAC CTATTGGTCT TACTGACATC CACTTTGCCT
      GAACGCAAAG ACTATCCGTG GATAACCAGA ATGACTGTAG GTGAAACGGA

1151  TTCTCTCCAC AGGTGTCCAC TCCCAGTTCA ATTACAGCTC TTAAAAGCTT
      AAGAGAGGTG TCCACAGGTG AGGGTCAAGT TAATGTCGAG AATTTTCGAA

                                          Met Glu Thr Leu Gln Phe Phe
1201  GGTACCGAGC TCGGATCTAT CGATGCCACC ATGGAGACCC TGCAGTTCTT
      CCATGGCTCG AGCCTAGATA GCTACGGTGG TACCTCTGGG ACGTCAAGAA
```

```
 1  Phe Phe Leu Phe Cys Cys Trp Lys Ala Ile Cys Cys Asn Ser Cys Glu Leu Thr
1251 CTTCCTGTTC TGCTGCTGGA AGGCCATCTG CTGCAACAGC TGCGAGCTGA
     GAAGGACAAG ACGACGACCT TCCGGTAGAC GACGTTGTCG ACGCTCGACT

 1  Thr Asn Ile Thr Ile Ala Ile Glu Lys Glu Glu Cys Arg Phe Cys Ile Ser
1301 CCAACATCAC CATCGCCATC GAGAAGGAGG AGTGCCGCTT CTGCATCAGC
     GGTTGTAGTG GTAGCGGTAG CTCTTCCTCC TCACGGCGAA GACGTAGTCG

 1  Ile Asn Thr Thr Trp Cys Ala Gly Trp Cys Trp Thr Arg Asp Leu Val Trp
1351 ATCAACACCA CCTGGTGCGC CGGCTACTGC TACACCCGCG ACCTGGTGTA
     TAGTTGTGGT GGACCACGCG GCCGATGACG ATGTGGGCGC TGGACCACAT

 1  Trp Lys Asp Pro Ala Arg Pro Lys Ile Gln Lys Thr Cys Thr Phe Lys Glu Leu
1401 CAAGGACCCC GCCCGCCCCA AGATCCAGAA GACCTGCACC TTCAAGGAGC
     GTTCCTGGGG CGGGCGGGGT TCTAGGTCTT CTGGACGTGG AAGTTCCTCG

 1  Leu Val Trp Glu Thr Val Arg Val Pro Gly Cys Ala His His Ala Asp Ser
1451 TGGTGTACGA GACGGTCCGG GTGCCCGGCT GCGCCCACCA CGCCGACAGC
     ACCACATGCT CTGCCAGGCC CACGGGCCGA CGCGGGTGGT GCGGCTGTCG

 1  Leu Trp Thr Tyr Pro Val Ala Thr Gln Cys His Cys Gly Lys Cys Asp Ser
1501 CTGTACACCT ACCCCGTGGC CACCCAGTGC CACTGCGGCA AGTGCGACAG
     GACATGTGGA TGGGGCACCG GTGGGTCACG GTGACGCCGT TCACGCTGTC

 1  Ser Asp Ser Thr Asp Cys Thr Val Arg Gly Leu Gly Pro Ser Trp Cys Ser Phe
1551 CGACAGCACC GACTGCACCG TGCGCGGCCT GGGCCCCAGC TACTGCAGCT
     GCTGTCGTGG CTGACGTGGC ACGCGCCGGA CCCGGGGTCG ATGACGTCGA

 1  Phe Gly Glu Met Lys
1601 TCGGCGAGAT GAAGGAGTAA CTCGAGACTA GAGGGCCCGT TTAAACCCGC
     AGCCGCTCTA CTTCCTCATT GAGCTCTGAT CTCCCGGGCA AATTTGGGCG

1651 TGATCAGCCT CGACTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTTGCCC
     ACTAGTCGGA GCTGACACGG AAGATCAACG GTCGGTAGAC AACAAACGGG

1701 CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCCTTT
     GAGGGGGCAC GGAAGGAACT GGGACCTTCC ACGGTGAGGG TGACAGGAAA

1751 CCTAATAAAA TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTCT
     GGATTATTTT ACTCCTTTAA CGTAGCGTAA CAGACTCATC CACAGTAAGA

1801 ATTCTGGGGG GTGGGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA
     TAAGACCCCC CACCCCACCC CGTCCTGTCG TTCCCCCTCC TAACCCTTCT

1851 CAATAGCAGG CATGCTGGGG ATGCGGTGGG CTCTATGGCT TCTGAGGCGG
     GTTATCGTCC GTACGACCCC TACGCCACCC GAGATACCGA AGACTCCGCC

1901 AAAGAACCAG CTGGGGCTCT AGGGGGTATC CCCACGCGCC CTGTAGCGGC
     TTTCTTGGTC GACCCCGAGA TCCCCCATAG GGGTGCGCGG GACATCGCCG

1951 GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT
     CGTAATTCGC GCCGCCCACA CCACCAATGC GCGTCGCACT GGCGATGTGA

2001 TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG
     ACGGTCGCGG GATCGCGGGC GAGGAAAGCG AAAGAAGGGA AGGAAAGAGC

2051 CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG CATCCCTTTA
     GGTGCAAGCG GCCGAAAGGG GCAGTTCGAG ATTTAGCCCC GTAGGGAAAT

2101 GGGTTCCGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA
     CCCAAGGCTA AATCACGAAA TGCCGTGGAG CTGGGGTTTT TTGAACTAAT

2151 GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG GTTTTTCGCC
     CCCACTACCA AGTGCATCAC CCGGTAGCGG GACTATCTGC CAAAAAGCGG

2201 CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAAACT
     GAAACTGCAA CCTCAGGTGC AAGAAATTAT CACCTGAGAA CAAGGTTTGA

2251 GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT
     CCTTGTTGTG AGTTGGGATA GAGCCAGATA AGAAAACTAA ATATTCCCTA

2301 TTTGGGGGATT TCGGCCTATT GGTTAAAAAA TGAGCTGATT TAACAAAAAT
     AAACCCCTAA AGCCGGATAA CCAATTTTTT ACTCGACTAA ATTGTTTTTA
```

```
2351  TTAACGCGAA TTAATTCTGT GGAATGTGTG TCAGTTAGGG TGTGGAAAGT
      AATTGCGCTT AATTAAGACA CCTTACACAC AGTCAATCCC ACACCTTTCA

2401  CCCCAGGCTC CCCAGGCAGG CAGAAGTATG CAAAGCATGC ATCTCAATTA
      GGGGTCCGAG GGGTCCGTCC GTCTTCATAC GTTTCGTACG TAGAGTTAAT

2451  GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA
      CAGTCGTTGG TCCACACCTT TCAGGGGTCC GAGGGGTCGT CCGTCTTCAT

2501  TGCAAAGCAT GCATCTCAAT TAGTCAGCAA CCATAGTCCC GCCCCTAACT
      ACGTTTCGTA CGTAGAGTTA ATCAGTCGTT GGTATCAGGG CGGGGATTGA

2551  CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA
      GGCGGGTAGG GCGGGGATTG AGGCGGGTCA AGGCGGGTAA GAGGCGGGGT

2601  TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCTGCCT
      ACCGACTGAT TAAAAAAAAT AAATACGTCT CCGGCTCCGG CGGAGACGGA

2651  CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG CCTAGGCTTT
      GACTCGATAA GGTCTTCATC ACTCCTCCGA AAAAACCTCC GGATCCGAAA

2701  TGCAAAAAGC TCCCGGGAGC TTGTATATCC ATTTTCGGAT CTGATCAGCA
      ACGTTTTTCG AGGGCCCTCG AACATATAGG TAAAAGCCTA GACTAGTCGT

2751  CGTGTTGACA ATTAATCATC GGCATAGTAT ATCGGCATAG TATAATACGA
      GCACAACTGT TAATTAGTAG CCGTATCATA TAGCCGTATC ATATTATGCT

2801  CAAGGTGAGG AACTAAACCA TGGCCAAGTT GACCAGTGCC GTTCCGGTGC
      GTTCCACTCC TTGATTTGGT ACCGGTTCAA CTGGTCACGG CAAGGCCACG

2851  TCACCGCGCG CGACGTCGCC GGAGCGGTCG AGTTCTGGAC CGACCGGCTC
      AGTGGCGCGC GCTGCAGCGG CCTCGCCAGC TCAAGACCTG GCTGGCCGAG

2901  GGGTTCTCCC GGGACTTCGT GGAGGACGAC TTCGCCGGTG TGGTCCGGGA
      CCCAAGAGGG CCCTGAAGCA CCTCCTGCTG AAGCGGCCAC ACCAGGCCCT

2951  CGACGTGACC CTGTTCATCA GCGCGGTCCA GGACCAGGTG GTGCCGGACA
      GCTGCACTGG GACAAGTAGT CGCGCCAGGT CCTGGTCCAC CACGGCCTGT

3001  ACACCCTGGC CTGGGTGTGG GTGCGCGGCC TGGACGAGCT GTACGCCGAG
      TGTGGGACCG GACCCACACC CACGCGCCGG ACCTGCTCGA CATGCGGCTC

3051  TGGTCGGAGG TCGTGTCCAC GAACTTCCGG GACGCCTCCG GGCCGGCCAT
      ACCAGCCTCC AGCACAGGTG CTTGAAGGCC CTGCGGAGGC CCGGCCGGTA

3101  GACCGAGATC GGCGAGCAGC CGTGGGGGCG GGAGTTCGCC CTGCGCGACC
      CTGGCTCTAG CCGCTCGTCG GCACCCCCGC CCTCAAGCGG GACGCGCTGG

3151  CGGCCGGCAA CTGCGTGCAC TTCGTGGCCG AGGAGCAGGA CTGACACGTG
      GCCGGCCGTT GACGCACGTG AAGCACCGGC TCCTCGTCCT GACTGTGCAC

3201  CTACGAGATT TCGATTCCAC CGCCGCCTTC TATGAAAGGT TGGGCTTCGG
      GATGCTCTAA AGCTAAGGTG GCGGCGGAAG ATACTTTCCA ACCCGAAGCC

3251  AATCGTTTTC CGGGACGCCG GCTGGATGAT CCTCCAGCGC GGGGATCTCA
      TTAGCAAAAG GCCCTGCGGC CGACCTACTA GGAGGTCGCG CCCCTAGAGT

3301  TGCTGGAGTT CTTCGCCCAC CCCAACTTGT TTATTGCAGC TTATAATGGT
      ACGACCTCAA GAAGCGGGTG GGGTTGAACA AATAACGTCG AATATTACCA

3351  TACAAATAAA GCAATAGCAT CACAAATTTC ACAAATAAAG CATTTTTTTC
      ATGTTTATTT CGTTATCGTA GTGTTTAAAG TGTTTATTTC GTAAAAAAAG

3401  ACTGCATTCT AGTTGTGGTT TGTCCAAACT CATCAATGTA TCTTATCATG
      TGACGTAAGA TCAACACCAA ACAGGTTTGA GTAGTTACAT AGAATAGTAC

3451  TCTGTATACC GTCGACCTCT AGCTAGAGCT TGGCGTAATC ATGGTCATAG
      AGACATATGG CAGCTGGAGA TCGATCTCGA ACCGCATTAG TACCAGTATC

3501  CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC ACAACATACG
      GACAAAGGAC ACACTTTAAC AATAGGCGAG TGTTAAGGTG TGTTGTATGC

3551  AGCCGGAAGC ATAAAGTGTA AAGCCTGGGG TGCCTAATGA GTGAGCTAAC
      TCGGCCTTCG TATTTCACAT TTCGGACCCC ACGGATTACT CACTCGATTG

3601  TCACATTAAT TGCGTTGCGC TCACTGCCCG CTTTCCAGTC GGGAAACCTG
      AGTGTAATTA ACGCAACGCG AGTGACGGGC GAAAGGTCAG CCCTTTGGAC
```

```
3651  TCGTGCCAGC TGCATTAATG AATCGGCCAA CGCGCGGGGA GAGGCGGTTT
      AGCACGGTCG ACGTAATTAC TTAGCCGGTT GCGCGCCCCT CTCCGCCAAA

3701  GCGTATTGGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
      CGCATAACCC GCGAGAAGGC GAAGGAGCGA GTGACTGAGC GACGCGAGCC

3751  TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
      AGCAAGCCGA CGCCGCTCGC CATAGTCGAG TGAGTTTCCG CCATTATGCC

3801  TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
      AATAGGTGTC TTAGTCCCCT ATTGCGTCCT TTCTTGTACA CTCGTTTTCC

3851  CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
      GGTCGTTTTC CGGTCCTTGG CATTTTTCCG GCGCAACGAC CGCAAAAAGG

3901  ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
      TATCCGAGGC GGGGGGACTG CTCGTAGTGT TTTTAGCTGC GAGTTCAGTC

3951  AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
      TCCACCGCTT TGGGCTGTCC TGATATTTCT ATGGTCCGCA AAGGGGGACC

4001  AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
      TTCGAGGGAG CACGCGAGAG GACAAGGCTG GGACGGCGAA TGGCCTATGG

4051  TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA ATGCTCACGC
      ACAGGCGGAA AGAGGGAAGC CCTTCGCACC GCGAAAGAGT TACGAGTGCG

4101  TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT
      ACATCCATAG AGTCAAGCCA CATCCAGCAA GCGAGGTTCG ACCCGACACA

4151  GCACGAACCC CCCGTTCAGC CCGACCGCTG CGCCTTATCC GGTAACTATC
      CGTGCTTGGG GGGCAAGTCG GGCTGGCGAC GCGGAATAGG CCATTGATAG

4201  GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC
      CAGAACTCAG GTTGGGCCAT TCTGTGCTGA ATAGCGGTGA CCGTCGTCGG

4251  ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT
      TGACCATTGT CCTAATCGTC TCGCTCCATA CATCCGCCAC GATGTCTCAA

4301  CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGGACA GTATTTGGTA
      GAACTTCACC ACCGGATTGA TGCCGATGTG ATCTTCCTGT CATAAACCAT

4351  TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT
      AGACGCGAGA CGACTTCGGT CAATGGAAGC CTTTTTCTCA ACCATCGAGA

4401  TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA
      ACTAGGCCGT TTGTTTGGTG GCGACCATCG CCACCAAAAA AACAAACGTT

4451  GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT CCTTTGATCT
      CGTCGTCTAA TGCGCGTCTT TTTTTCCTAG AGTTCTTCTA GGAAACTAGA

4501  TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT
      AAAGATGCCC CAGACTGCGA GTCACCTTGC TTTTGAGTGC AATTCCCTAA

4551  TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA
      AACCAGTACT CTAATAGTTT TTCCTAGAAG TGGATCTAGG AAAATTTAAT

4601  AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG
      TTTTACTTCA AAATTTAGTT AGATTTCATA TATACTCATT TGAACCAGAC

4651  ACAGTTACCA ATGCTTAATC AGTGAGGCAC CTATCTCAGC GATCTGTCTA
      TGTCAATGGT TACGAATTAG TCACTCCGTG GATAGAGTCG CTAGACAGAT

4701  TTTCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT
      AAAGCAAGTA GGTATCAACG GACTGAGGGG CAGCACATCT ATTGATGCTA

4751  ACGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA CCGCGAGACC
      TGCCCTCCCG AATGGTAGAC CGGGGTCACG ACGTTACTAT GGCGCTCTGG

4801  CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG
      GTGCGAGTGG CCGAGGTCTA AATAGTCGTT ATTTGGTCGG TCGGCCTTCC

4851  GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA TCCAGTCTAT
      CGGCTCGCGT CTTCACCAGG ACGTTGAAAT AGGCGGAGGT AGGTCAGATA

4901  TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC
      ATTAACAACG GCCCTTCGAT CTCATTCATC AAGCGGTCAA TTATCAAACG
```

```
4951  GCAACGTTGT TGCCATTGCT ACAGGCATCG TGGTGTCACG CTCGTCGTTT
      CGTTGCAACA ACGGTAACGA TGTCCGTAGC ACCACAGTGC GAGCAGCAAA

5001  GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG
      CCATACCGAA GTAAGTCGAG GCCAAGGGTT GCTAGTTCCG CTCAATGTAC

5051  ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT CCTCCGATCG
      TAGGGGGTAC AACACGTTTT TTCGCCAATC GAGGAAGCCA GGAGGCTAGC

5101  TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA
      AACAGTCTTC ATTCAACCGG CGTCACAATA GTGAGTACCA ATACCGTCGT

5151  CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC
      GACGTATTAA GAGAATGACA GTACGGTAGG CATTCTACGA AAAGACACTG

5201  TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA
      ACCACTCATG AGTTGGTTCA GTAAGACTCT TATCACATAC GCCGCTGGCT

5251  GTTGCTCTTG CCCGGCGTCA ATACGGGATA ATACCGCGCC ACATAGCAGA
      CAACGAGAAC GGGCCGCAGT TATGCCCTAT TATGGCGCGG TGTATCGTCT

5301  ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC
      TGAAATTTTC ACGAGTAGTA ACCTTTTGCA AGAAGCCCCG CTTTTGAGAG

5351  AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC ACTCGTGCAC
      TTCCTAGAAT GGCGACAACT CTAGGTCAAG CTACATTGGG TGAGCACGTG

5401  CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA
      GGTTGACTAG AAGTCGTAGA AAATGAAAGT GGTCGCAAAG ACCCACTCGT

5451  AAAACAGGAA GGCAAAATGC CGCAAAAAAG GGAATAAGGG CGACACGGAA
      TTTTGTCCTT CCGTTTTACG GCGTTTTTTC CCTTATTCCC GCTGTGCCTT

5501  ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC
      TACAACTTAT GAGTATGAGA AGGAAAAAGT TATAATAACT TCGTAAATAG

5551  AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT
      TCCCAATAAC AGAGTACTCG CCTATGTATA AACTTACATA AATCTTTTTA

5601  AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT
      TTTGTTTATC CCCAAGGCGC GTGTAAAGGG GCTTTTCACG GTGGACTGCA

5651  C
      G
```

FIGURE 4 (p. 1/5)

```
   1  GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC
      CTGCCTAGCC CTCTAGAGGG CTAGGGGATA CCAGCTGAGA GTCATGTTAG

  51  TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
      ACGAGACTAC GGCGTATCAA TTCGGTCATA GACGAGGGAC GAACACACAA

 101  GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
      CCTCCAGCGA CTCATCACGC GCTCGTTTTA AATTCGATGT TGTTCCGTTC

 151  GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG
      CGAACTGGCT GTTAACGTAC TTCTTAGACG AATCCCAATC CGCAAAACGC

 201  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT GATTATTGAC
      GACGAAGCGC TACATGCCCG GTCTATATGC GCAACTGTAA CTAATAACTG

 251  TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
      ATCAATAATT ATCATTAGTT AATGCCCCAG TAATCAAGTA TCGGGTATAT

 301  TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG
      ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC

 351  CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
      GGGTTGCTGG GGGCGGGTAA CTGCAGTTAT TACTGCATAC AAGGGTATCA

 401  AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAC TATTTACGGT
      TTGCGGTTAT CCCTGAAAGG TAACTGCAGT TACCCACCTG ATAAATGCCA

 451  AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC
      TTTGACGGGT GAACCGTCAT GTAGTTCACA TAGTATACGG TTCATGCGGG

 501  CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA
      GGATAACTGC AGTTACTGCC ATTTACCGGG CGGACCGTAA TACGGGTCAT

 551  CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA
      GTACTGGAAT ACCCTGAAAG GATGAACCGT CATGTAGATG CATAATCAGT

 601  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA
      AGCGATAATG GTACCACTAC GCCAAAACCG TCATGTAGTT ACCCGCACCT

 651  TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
      ATCGCCAAAC TGAGTGCCCC TAAAGGTTCA GAGGTGGGGT AACTGCAGTT

 701  TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA
      ACCCTCAAAC AAAACCGTGG TTTTAGTTGC CCTGAAAGGT TTTACAGCAT

 751  ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG
      TGTTGAGGCG GGGTAACTGC GTTTACCCGC CATCCGCACA TGCCACCCTC

 801  GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG
      CAGATATATT CGTCTCGAGA GACCGATTGA TCTCTTGGGT GACGAATGAC

 851  GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
      CGAATAGCTT TAATTATGCT GAGTGATATC CCTCTGGGTT CGACCGATCG

 901  TTATTGCGGT AGTTTATCAC AGTTAAATTG CTAACGCAGT CAGTGCTTCT
      AATAACGCCA TCAAATAGTG TCAATTTAAC GATTGCGTCA GTCACGAAGA

 951  GACACAACAG TCTCGAACTT AAGCTGCAGT GACTCTCTTA AGGTAGCCTT
      CTGTGTTGTC AGAGCTTGAA TTCGACGTCA CTGAGAGAAT TCCATCGGAA

1001  GCAGAAGTTG GTCGTGAGGC ACTGGGCAGG TAAGTATCAA GGTTACAAGA
      CGTCTTCAAC CAGCACTCCG TGACCCGTCC ATTCATAGTT CCAATGTTCT

1051  CAGGTTTAAG GAGACCAATA GAAACTGGGC TTGTCGAGAC AGAGAAGACT
      GTCCAAATTC CTCTGGTTAT CTTTGACCCG AACAGCTCTG TCTCTTCTGA

1101  CTTGCGTTTC TGATAGGCAC CTATTGGTCT TACTGACATC CACTTTGCCT
      GAACGCAAAG ACTATCCGTG GATAACCAGA ATGACTGTAG GTGAAACGGA

1151  TTCTCTCCAC AGGTGTCCAC TCCCAGTTCA ATTACAGCTC TTAAAAGCTT
      AAGAGAGGTG TCCACAGGTG AGGGTCAAGT TAATGTCGAG AATTTTCGAA

 *1                                    Met  Asp  Tyr  Tyr  Arg  Lys  Tyr  Ala  Ala

1201  GGTACCGAGC TCGGATCCGC CACCATGGAC TACTACCGCA AGTACGCCGC
      CCATGGCTCG AGCCTAGGCG GTGGTACCTG ATGATGGCGT TCATGCGGCG
```

```
 1  Ala Ile Phe Leu Val Thr Leu Ser Val Phe Leu His Val Leu His Ser Ala Asn
1251 CATCTTCCTG GTGACCCTGA GCGTGTTCCT GCACGTGCTG CACAGCGCCA
     GTAGAAGGAC CACTGGGACT CGCACAAGGA CGTGCACGAC GTGTCGCGGT

 1  Asn Ile Thr Val Asn Ile Thr Val Ala Pro Asp Val Gln Asp Cys Pro Glu
1301 ACATCACCGT TAACATCACC GTGGCCCCCG ACGTGCAGGA CTGCCCCGAG
     TGTAGTGGCA ATTGTAGTGG CACCGGGGGC TGCACGTCCT GACGGGGCTC

 1  Cys Thr Leu Gln Glu Asn Pro Phe Phe Ser Gln Pro Gly Ala Pro Ile Leu
1351 TGCACCCTGC AGGAGAACCC CTTCTTCAGC CAGCCCGGCG CCCCCATCCT
     ACGTGGGACG TCCTCTTGGG GAAGAAGTCG GTCGGGCCGC GGGGGTAGGA

 1  Leu Gln Cys Met Gly Cys Cys Phe Ser Arg Ala Trp Pro Thr Pro Leu Arg Ser
1401 GCAGTGCATG GGCTGCTGCT TCAGCCGCGC CTACCCCACC CCCCTGCGCA
     CGTCACGTAC CCGACGACGA AGTCGGCGCG GATGGGGTGG GGGGACGCGT

 1  Ser Lys Lys Thr Met Leu Val Gln Lys Asn Val Thr Ser Glu Ser Thr Cys
1451 GCAAGAAGAC CATGCTGGTG CAGAAGAACG TGACCAGCGA GAGCACCTGC
     CGTTCTTCTG GTACGACCAC GTCTTCTTGC ACTGGTCGCT CTCGTGGACG

 1  Cys Val Ala Lys Ser Tyr Asn Arg Val Thr Val Met Gly Gly Phe Lys Val
1501 TGCGTGGCCA AGAGCTACAA CCGCGTGACC GTGATGGGCG GCTTCAAGGT
     ACGCACCGGT TCTCGATGTT GGCGCACTGG CACTACCCGC CGAAGTTCCA

 1  Val Glu Asn His Thr Ala Cys His Cys Ser Thr Cys Tyr Tyr His Lys
1551 GGAGAACCAC ACCGCCTGCC ACTGCAGCAC CTGCTACTAC CACAAGAGCT
     CCTCTTGGTG TGGCGGACGG TGACGTCGTG GACGATGATG GTGTTCTCGA

1601 AATCTAGAGG GCCCGTTTAA ACCCGCTGAT CAGCCTCGAC TGTGCCTTCT
     TTAGATCTCC CGGGCAAATT TGGGCGACTA GTCGGAGCTG ACACGGAAGA

1651 AGTTGCCAGC CATCTGTTGT TTGCCCCTCC CCCGTGCCTT CCTTGACCCT
     TCAACGGTCG GTAGACAACA AACGGGGAGG GGGCACGGAA GGAACTGGGA

1701 GGAAGGTGCC ACTCCCACTG TCCTTTCCTA ATAAAATGAG GAAATTGCAT
     CCTTCCACGG TGAGGGTGAC AGGAAAGGAT TATTTTACTC CTTTAACGTA

1751 CGCATTGTCT GAGTAGGTGT CATTCTATTC TGGGGGGTGG GGTGGGGCAG
     GCGTAACAGA CTCATCCACA GTAAGATAAG ACCCCCCACC CCACCCCGTC

1801 GACAGCAAGG GGGAGGATTG GGAAGACAAT AGCAGGCATG CTGGGGATGC
     CTGTCGTTCC CCCTCCTAAC CCTTCTGTTA TCGTCCGTAC GACCCCTACG

1851 GGTGGGCTCT ATGGCTTCTG AGGCGGAAAG AACCAGCTGG GGCTCTAGGG
     CCACCCGAGA TACCGAAGAC TCCGCCTTTC TTGGTCGACC CCGAGATCCC

1901 GGTATCCCCA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG
     CCATAGGGGT GCGCGGGACA TCGCCGCGTA ATTCGCGCCG CCCACACCAC

1951 GTTACGCGCA GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC
     CAATGCGCGT CGCACTGGCG ATGTGAACGG TCGCGGGATC GCGGGCGAGG

2001 TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC
     AAAGCGAAAG AAGGGAAGGA AAGAGCGGTG CAAGCGGCCG AAAGGGGCAG

2051 AAGCTCTAAA TCGGGGCATC CCTTTAGGGT TCCGATTTAG TGCTTTACGG
     TTCGAGATTT AGCCCCGTAG GGAAATCCCA AGGCTAAATC ACGAAATGCC

2101 CACCTCGACC CCAAAAAACT TGATTAGGGT GATGGTTCAC GTAGTGGGCC
     GTGGAGCTGG GGTTTTTTGA ACTAATCCCA CTACCAAGTG CATCACCCGG

2151 ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT
     TAGCGGGACT ATCTGCCAAA AAGCGGGAAA CTGCAACCTC AGGTGCAAGA

2201 TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCTATCTCG
     AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC

2251 GTCTATTCTT TTGATTTATA AGGGATTTTG GGGATTTCGG CCTATTGGTT
     CAGATAAGAA AACTAAATAT TCCCTAAAAC CCCTAAAGCC GGATAACCAA

2301 AAAAAATGAG CTGATTTAAC AAAAATTTAA CGCGAATTAA TTCTGTGGAA
     TTTTTTACTC GACTAAATTG TTTTTAAATT GCGCTTAATT AAGACACCTT

2351 TGTGTGTCAG TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GGCAGGCAGA
     ACACACAGTC AATCCCACAC CTTTCAGGGG TCCGAGGGGT CCGTCCGTCT
```

```
2401  AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT GTGGAAAGTC
      TCATACGTTT CGTACGTAGA GTTAATCAGT CGTTGGTCCA CACCTTTCAG

2451  CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT
      GGGTCCGAGG GGTCGTCCGT CTTCATACGT TTCGTACGTA GAGTTAATCA

2501  CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG
      GTCGTTGGTA TCAGGGCGGG GATTGAGGCG GGTAGGGCGG GGATTGAGGC

2551  CCCAGTTCCG CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA
      GGGTCAAGGC GGGTAAGAGG CGGGGTACCG ACTGATTAAA AAAAATAAAT

2601  TGCAGAGGCC GAGGCCGCCT CTGCCTCTGA GCTATTCCAG AAGTAGTGAG
      ACGTCTCCGG CTCCGGCGGA GACGGAGACT CGATAAGGTC TTCATCACTC

2651  GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA AAAAGCTCCC GGGAGCTTGT
      CTCCGAAAAA ACCTCCGGAT CCGAAAACGT TTTTCGAGGG CCCTCGAACA

2701  ATATCCATTT TCGGATCTGA TCAGCACGTG ATGAAAAAGC CTGAACTCAC
      TATAGGTAAA AGCCTAGACT AGTCGTGCAC TACTTTTTCG GACTTGAGTG

2751  CGCGACGTCT GTCGAGAAGT TTCTGATCGA AAAGTTCGAC AGCGTCTCCG
      GCGCTGCAGA CAGCTCTTCA AAGACTAGCT TTTCAAGCTG TCGCAGAGGC

2801  ACCTGATGCA GCTCTCGGAG GGCGAAGAAT CTCGTGCTTT CAGCTTCGAT
      TGGACTACGT CGAGAGCCTC CCGCTTCTTA GAGCACGAAA GTCGAAGCTA

2851  GTAGGAGGGC GTGGATATGT CCTGCGGGTA AATAGCTGCG CCGATGGTTT
      CATCCTCCCG CACCTATACA GGACGCCCAT TTATCGACGC GGCTACCAAA

2901  CTACAAAGAT CGTTATGTTT ATCGGCACTT TGCATCGGCC GCGCTCCCGA
      GATGTTTCTA GCAATACAAA TAGCCGTGAA ACGTAGCCGG CGCGAGGGCT

2951  TTCCGGAAGT GCTTGACATT GGGGAATTCA GCGAGAGCCT GACCTATTGC
      AAGGCCTTCA CGAACTGTAA CCCCTTAAGT CGCTCTCGGA CTGGATAACG

3001  ATCTCCCGCC GTGCACAGGG TGTCACGTTG CAAGACCTGC CTGAAACCGA
      TAGAGGGCGG CACGTGTCCC ACAGTGCAAC GTTCTGGACG GACTTTGGCT

3051  ACTGCCCGCT GTTCTGCAGC CGGTCGCGGA GGCCATGGAT GCGATCGCTG
      TGACGGGCGA CAAGACGTCG GCCAGCGCCT CCGGTACCTA CGCTAGCGAC

3101  CGGCCGATCT TAGCCAGACG AGCGGGTTCG GCCCATTCGG ACCGCAAGGA
      GCCGGCTAGA ATCGGTCTGC TCGCCCAAGC CGGGTAAGCC TGGCGTTCCT

3151  ATCGGTCAAT ACACTACATG GCGTGATTTC ATATGCGCGA TTGCTGATCC
      TAGCCAGTTA TGTGATGTAC CGCACTAAAG TATACGCGCT AACGACTAGG

3201  CCATGTGTAT CACTGGCAAA CTGTGATGGA CGACACCGTC AGTGCGTCCG
      GGTACACATA GTGACCGTTT GACACTACCT GCTGTGGCAG TCACGCAGGC

3251  TCGCGCAGGC TCTCGATGAG CTGATGCTTT GGGCCGAGGA CTGCCCCGAA
      AGCGCGTCCG AGAGCTACTC GACTACGAAA CCCGGCTCCT GACGGGGCTT

3301  GTCCGGCACC TCGTGCACGC GGATTTCGGC TCCAACAATG TCCTGACGGA
      CAGGCCGTGG AGCACGTGCG CCTAAAGCCG AGGTTGTTAC AGGACTGCCT

3351  CAATGGCCGC ATAACAGCGG TCATTGACTG GAGCGAGGCG ATGTTCGGGG
      GTTACCGGCG TATTGTCGCC AGTAACTGAC CTCGCTCCGC TACAAGCCCC

3401  ATTCCCAATA CGAGGTCGCC AACATCTTCT TCTGGAGGCC GTGGTTGGCT
      TAAGGGTTAT GCTCCAGCGG TTGTAGAAGA AGACCTCCGG CACCAACCGA

3451  TGTATGGAGC AGCAGACGCG CTACTTCGAG CGGAGGCATC CGGAGCTTGC
      ACATACCTCG TCGTCTGCGC GATGAAGCTC GCCTCCGTAG GCCTCGAACG

3501  AGGATCGCCG CGGCTCCGGG CGTATATGCT CCGCATTGGT CTTGACCAAC
      TCCTAGCGGC GCCGAGGCCC GCATATACGA GGCGTAACCA GAACTGGTTG

3551  TCTATCAGAG CTTGGTTGAC GGCAATTTCG ATGATGCAGC TTGGGCGCAG
      AGATAGTCTC GAACCAACTG CCGTTAAAGC TACTACGTCG AACCCGCGTC

3601  GGTCGATGCG ACGCAATCGT CCGATCCGGA GCCGGGACTG TCGGGCGTAC
      CCAGCTACGC TGCGTTAGCA GGCTAGGCCT CGGCCCTGAC AGCCGCATG

3651  ACAAATCGCC CGCAGAAGCG CGGCCGTCTG GACCGATGGC TGTGTAGAAG
      TGTTTAGCGG GCGTCTTCGC GCCGGCAGAC CTGGCTACCG ACACATCTTC
```

```
3701  TACTCGCCGA TAGTGGAAAC CGACGCCCCA GCACTCGTCC GAGGGCAAAG
      ATGAGCGGCT ATCACCTTTG GCTGCGGGGT CGTGAGCAGG CTCCCGTTTC

3751  GAATAGCACG TGCTACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG
      CTTATCGTGC ACGATGCTCT AAAGCTAAGG TGGCGGCGGA AGATACTTTC

3801  GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC
      CAACCCGAAG CCTTAGCAAA AGGCCCTGCG GCCGACCTAC TAGGAGGTCG

3851  GCGGGGATCT CATGCTGGAG TTCTTCGCCC ACCCCAACTT GTTTATTGCA
      CGCCCCTAGA GTACGACCTC AAGAAGCGGG TGGGGTTGAA CAAATAACGT

3901  GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT TCACAAATAA
      CGAATATTAC CAATGTTTAT TTCGTTATCG TAGTGTTTAA AGTGTTTATT

3951  AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG
      TCGTAAAAAA AGTGACGTAA GATCAACACC AAACAGGTTT GAGTAGTTAC

4001  TATCTTATCA TGTCTGTATA CCGTCGACCT CTAGCTAGAG CTTGGCGTAA
      ATAGAATAGT ACAGACATAT GGCAGCTGGA GATCGATCTC GAACCGCATT

4051  TCATGGTCAT AGCTGTTTCC TGTGTGAAAT TGTTATCCGC TCACAATTCC
      AGTACCAGTA TCGACAAAGG ACACACTTTA ACAATAGGCG AGTGTTAAGG

4101  ACACAACATA CGAGCCGGAA GCATAAAGTG TAAAGCCTGG GGTGCCTAAT
      TGTGTTGTAT GCTCGGCCTT CGTATTTCAC ATTTCGGACC CCACGGATTA

4151  GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCCAG
      CTCACTCGAT TGAGTGTAAT TAACGCAACG CGAGTGACGG GCGAAAGGTC

4201  TCGGGAAACC TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGCGGG
      AGCCCTTTGG ACAGCACGGT CGACGTAATT ACTTAGCCGG TTGCGCGCCC

4251  GAGAGGCGGT TTGCGTATTG GGCGCTCTTC CGCTTCCTCG CTCACTGACT
      CTCTCCGCCA AACGCATAAC CCGCGAGAAG GCGAAGGAGC GAGTGACTGA

4301  CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG
      GCGACGCGAG CCAGCAAGCC GACGCCGCTC GCCATAGTCG AGTGAGTTTC

4351  GCGGTAATAC GGTTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT
      CGCCATTATG CCAATAGGTG TCTTAGTCCC CTATTGCGTC CTTTCTTGTA

4401  GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC
      CACTCGTTTT CCGGTCGTTT TCCGGTCCTT GGCATTTTTC CGGCGCAACG

4451  TGGCGTTTTT CCATAGGCTC CGCCCCCCTG ACGAGCATCA CAAAAATCGA
      ACCGCAAAAA GGTATCCGAG GCGGGGGGAC TGCTCGTAGT GTTTTTAGCT

4501  CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC
      GCGAGTTCAG TCTCCACCGC TTTGGGCTGT CCTGATATTT CTATGGTCCG

4551  GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC
      CAAAGGGGGA CCTTCGAGGG AGCACGCGAG AGGACAAGGC TGGGACGGCG

4601  TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT
      AATGGCCTAT GGACAGGCGG AAAGAGGGAA GCCCTTCGCA CCGCGAAAGA

4651  CAATGCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA
      GTTACGAGTG CGACATCCAT AGAGTCAAGC CACATCCAGC AAGCGAGGTT

4701  GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT
      CGACCCGACA CACGTGCTTG GGGGGCAAGT CGGGCTGGCG ACGCGGAATA

4751  CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATCGCCA
      GGCCATTGAT AGCAGAACTC AGGTTGGGCC ATTCTGTGCT GAATAGCGGT

4801  CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTAGGCGG
      GACCGTCGTC GGTGACCATT GTCCTAATCG TCTCGCTCCA TACATCCGCC

4851  TGCTACAGAG TTCTTGAAGT GGTGGCCTAA CTACGGCTAC ACTAGAAGGA
      ACGATGTCTC AAGAACTTCA CCACCGGATT GATGCCGATG TGATCTTCCT

4901  CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA
      GTCATAAACC ATAGACGCGA GACGACTTCG GTCAATGGAA GCCTTTTTCT

4951  GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT
      CAACCATCGA GAACTAGGCC GTTTGTTTGG TGGCGACCAT CGCCACCAAA
```

```
5001  TTTTGTTTGC AAGCAGCAGA TTACGCGCAG AAAAAAAGGA TCTCAAGAAG
      AAAACAAACG TTCGTCGTCT AATGCGCGTC TTTTTTTCCT AGAGTTCTTC

5051  ATCCTTTGAT CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACTCA
      TAGGAAACTA GAAAAGATGC CCCAGACTGC GAGTCACCTT GCTTTTGAGT

5101  CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT
      GCAATTCCCT AAAACCAGTA CTCTAATAGT TTTTCCTAGA AGTGGATCTA

5151  CCTTTTAAAT TAAAAATGAA GTTTTAAATC AATCTAAAGT ATATATGAGT
      GGAAAATTTA ATTTTTACTT CAAAATTTAG TTAGATTTCA TATATACTCA

5201  AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA
      TTTGAACCAG ACTGTCAATG GTTACGAATT AGTCACTCCG TGGATAGAGT

5251  GCGATCTGTC TATTTCGTTC ATCCATAGTT GCCTGACTCC CCGTCGTGTA
      CGCTAGACAG ATAAAGCAAG TAGGTATCAA CGGACTGAGG GGCAGCACAT

5301  GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA
      CTATTGATGC TATGCCCTCC CGAATGGTAG ACCGGGGTCA CGACGTTACT

5351  TACCGCGAGA CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCAG
      ATGGCGCTCT GGGTGCGAGT GGCCGAGGTC TAAATAGTCG TTATTTGGTC

5401  CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT CCTGCAACTT TATCCGCCTC
      GGTCGGCCTT CCCGGCTCGC GTCTTCACCA GGACGTTGAA ATAGGCGGAG

5451  CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAAGT AGTTCGCCAG
      GTAGGTCAGA TAATTAACAA CGGCCCTTCG ATCTCATTCA TCAAGCGGTC

5501  TTAATAGTTT GCGCAACGTT GTTGCCATTG CTACAGGCAT CGTGGTGTCA
      AATTATCAAA CGCGTTGCAA CAACGGTAAC GATGTCCGTA GCACCACAGT

5551  CGCTCGTCGT TTGGTATGGC TTCATTCAGC TCCGGTTCCC AACGATCAAG
      GCGAGCAGCA AACCATACCG AAGTAAGTCG AGGCCAAGGG TTGCTAGTTC

5601  GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG
      CGCTCAATGT ACTAGGGGGT ACAACACGTT TTTTCGCCAA TCGAGGAAGC

5651  GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG CCGCAGTGTT ATCACTCATG
      CAGGAGGCTA GCAACAGTCT TCATTCAACC GGCGTCACAA TAGTGAGTAC

5701  GTTATGGCAG CACTGCATAA TTCTCTTACT GTCATGCCAT CCGTAAGATG
      CAATACCGTC GTGACGTATT AAGAGAATGA CAGTACGGTA GGCATTCTAC

5751  CTTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGTA
      GAAAAGACAC TGACCACTCA TGAGTTGGTT CAGTAAGACT CTTATCACAT

5801  TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA TAATACCGCG
      ACGCCGCTGG CTCAACGAGA ACGGGCCGCA GTTATGCCCT ATTATGGCGC

5851  CCACATAGCA GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG
      GGTGTATCGT CTTGAAATTT TCACGAGTAG TAACCTTTTG CAAGAAGCCC

5901  GCGAAAACTC TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC
      CGCTTTTGAG AGTTCCTAGA ATGGCGACAA CTCTAGGTCA AGCTACATTG

5951  CCACTCGTGC ACCCAACTGA TCTTCAGCAT CTTTTACTTT CACCAGCGTT
      GGTGAGCACG TGGGTTGACT AGAAGTCGTA GAAAATGAAA GTGGTCGCAA

6001  TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA AGGGAATAAG
      AGACCCACTC GTTTTTGTCC TTCCGTTTTA CGGCGTTTTT TCCCTTATTC

6051  GGCGACACGG AAATGTTGAA TACTCATACT CTTCCTTTTT CAATATTATT
      CCGCTGTGCC TTTACAACTT ATGAGTATGA GAAGGAAAAA GTTATAATAA

6101  GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT
      CTTCGTAAAT AGTCCCAATA ACAGAGTACT CGCCTATGTA TAAACTTACA

6151  ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT
      TAAATCTTTT TATTTGTTTA TCCCCAAGGC GCGTGTAAAG GGGCTTTTCA

6201  GCCACCTGAC GTC
      CGGTGGACTG CAG
```